

Louisiana State University LSU Digital Commons

LSU Master's Theses

Graduate School

2014

Analyses of Online Monitoring Signals for a GMAW Process Before and After Improvement

Aniruddha Joshi

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_theses



Part of the [Mechanical Engineering Commons](#)

Recommended Citation

Joshi, Aniruddha, "Analyses of Online Monitoring Signals for a GMAW Process Before and After Improvement" (2014). *LSU Master's Theses*. 4008.

https://digitalcommons.lsu.edu/gradschool_theses/4008

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

ANALYSES OF ONLINE MONITORING SIGNALS FOR A GMAW PROCESS BEFORE AND AFTER IMPROVEMENT

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Science in Industrial Engineering

in

The Department of Mechanical and Industrial Engineering

by
Aniruddha Joshi
B.S., Georgia Institute of Technology, 2012
December 2014

ACKNOWLEDGMENTS

I would like to sincerely thank my advisor, Dr. T. Warren Liao for all the help which he has given me during my time in this program. This research was only possible through his invaluable guidance and direction. I would also like to thank Dr. Marcio de Queiroz, Dr. Bin Li and Dr. Muhammad Wahab who have taken the time to serve as a part of my Masters committee and have provided useful insights throughout the research process. Furthermore, I would like to thank Dr. Lampros Kompotiatis for his valuable insight in this research project.

I would also like to thank my family and friends for their support.

TABLE OF CONTENTS

ACKNOWLEDGMENTS.....	ii
ABSTRACT.....	iv
CHAPTER 1 – INTRODUCTION	1
1.1 Background of Problem	2
CHAPTER 2 - LITERATURE REVIEW	3
CHAPTER 3- RESEARCH TOPIC AND GOAL.....	10
CHAPTER 4 – EXPLORATORY DATA ANALYSIS	11
CHAPTER 5 – METHODOLOGY	18
5.1 Feature Extraction	19
5.1.1 Probability Distributions.....	19
5.1.2 Energy Spectra	20
5.1.3 Wavelet Transform	21
5.1.4 Other Time Domain Features.....	25
5.2 Instance Selection	27
5.3 Classification Model	29
5.3.1 Neural network modeling	30
5.3.2 Fuzzy K-NN modeling.....	34
5.4 Feature Selection.....	36
5.4.1 Wrapper Approach.....	37
5.4.2 Ant Colony Optimization.....	39
5.4.3 ACO-S Algorithm.....	41
CHAPTER 6 – RESULTS AND DISCUSSION	44
6.1 Classification Model	44
6.2 Feature Selection.....	47
6.3 Ranking the features	50
6.4 Effects of adjusting the window size	53
6.5 Application of the feature selection results to entire dataset	56
CHAPTER 7 - SUMMARY AND CONCLUSION.....	61
REFERENCES	64
APPENDIX.....	70
VITA	71

ABSTRACT

The ability to detect the onset of welding instability is a very powerful tool in welding process monitoring and control. Toward this goal, this study investigates a gas metal arc welding (GMAW) process by analysing online monitoring signals. Two separate data sets are obtained from the process, which correspond to (a) a stable process after improvement and (b) a relatively unstable process which tends to exhibit spatter and poor weld bead geometry. Voltage, current, wire-feeding speed and line speed signals for both data sets are analysed and features are extracted from the raw signals using different signal processing techniques. Specifically, phase diagrams, signal distributions, Fast Fourier transform (FFT) and Wavelet Transform methodologies are implemented. The process parameters differ for the data corresponding to the stable and unstable processes rendering the two data sets incomparable. As such, an overlapping region of parameters is selected and this data is used to develop a multi-layer neural network model. The model uses the features extracted to distinguish between the two data sets under the similar input conditions. The trained model is then used to classify data as being from a stable process or an unstable process. Finally, an ant colony optimization algorithm is used to select the optimal subset of features for the classification model. For this task, fuzzy k-nearest neighbor algorithm is used as the classifier instead due to the computational simplicity. The results indicate that more than one single feature is able to yield 100% classification accuracy alone. A way to rank those features is discussed. Moreover, the effect of window size is also investigated.

CHAPTER 1 – INTRODUCTION

Gas Metal Arc Welding (GMAW) is a major welding process used in various engineering applications. GMAW involves the use of a high energy electrical arc shielded by a fed gas to coalesce two or more pieces of metal. The shielding gas, usually either an argon and carbon dioxide mixture or pure carbon dioxide gas, serves to keep oxygen and other undesirable gases away from the welding area. Continuous GMAW welding has been widely used in automated manufacturing processes.

The GMAW process can be subdivided according to the metal transfer mode. The three major metal transfer modes operating in this process are short circuit, globular, and spray transfer. The metal transfer mode describes the way in which molten metal detaches from the electrode and deposits onto the work piece. CO₂ based GMAW, which is studied in this research, is for the most part capable of sustaining short circuit and globular transfer. This is due in part to the effects of the shielding gases on droplet formation. The GMAW process is also categorized as either Metal Inert Gas or Metal Active Gas welding depending on the shielding gas used. Processes using carbon dioxide gas, which is a semi-inert gas, belong to the Metal Active Gas (MAG) category.

One major problem associated with continuous GMAW is the incidence of weld spatter. Weld spatter refers to small droplets of weld material which are deposited haphazardly onto the welding surface or elsewhere. Spatter may or may not be accompanied by uneven deposition of weld material on the work piece. The prevalence of spatter and/or irregular deposition defines an unstable welding process. Operating in unstable conditions may result in several detriments. Firstly, irregularities in the weld bead due to uneven deposition may lead to poor quality finished welds. These translate into poorer structural integrity of the final product. Secondly, spatter may lead to increase in wasted material. Poorly welded material may be discarded due to excessive

spatter. As such spatter is undesirable in GMAW operations. To this end, the continuous GMAW processes in industry are often monitored in real time. Real time online weld quality monitoring is increasingly beneficial to industry as it prevents waste in production if welds are seen to be poor down the line. There are great financial advantages possible through real time monitoring (Wu et al. 2001).

1.1 Background of Problem

The data used in this study was obtained from a continuous MAG welding process in an industrial spiral pipe manufacturing facility. The process before improvement has a high tendency of process instability and the onset of weld spatter during operation. The process is monitored through four signals, namely the voltage, current, wire feeding speed and line speed. The nature of the signals will be explained in further detail in the following chapters. The output quality of the welding process is monitored automatically, but not recorded and consequently there is no quantitative measurement of the resulting weld available in real time.

The data obtained spanned over a time frame of over a year. During this time period enhancements were made to the process which resulted in significant improvements in process operation. The process before the improvement was seen to result in abundant spatter and poor quality welds. Conversely, the process after the improvement (primarily in the torch but details are unknown due to its proprietary nature) was seen to exhibit a very stable process with very minimal spatter and defects. As such, although there is no direct quality measure available, the general stability of the process before and after improvement is known.

CHAPTER 2 - LITERATURE REVIEW

Owing to the significant industrial benefits of the ability to determine welding process stability, there has been much effort in developing methods to accomplish this task. Online welding process and weld quality monitoring has been attempted using various sensing techniques. Infrared, ultrasound, visual, acoustic emission and arc sensor signals are a few of the tools used to monitor the welding process during production. Using any of these techniques generally involves analyses of the signals and modeling with the input-output data.

Acoustic emission has been an effective tool in instability recognition due to distinct sound patterns made by spatter (Grad et al. 2004). The sound made by spatter may even be detected by the human ear due to the sudden explosion of the wire tip commonly present during the occurrence of spatter. Visual systems assessing the resulting weld bead may experience disturbance of the sensor system due to the intense amounts of light dissipated by the arc (Chu et al. 2004). Visual systems have the advantage of directly observing any weld spatters or weld bead geometry defects. Aside from arc sensor signals, most of the techniques mentioned require an additional sensor attached to or near the torch. This may be difficult in some cases due to the minimal torch clearance maintained in many GMAW processes (Quinn et al. 1999). Arc sensor signals, specifically current and voltage, may be of more interest as they provide a direct and reliable way of observing the behavior of the arc. Arc sensing has been used as an online monitoring technique by Quinn et al. (1999) in an effort to detect flaws in an automatic GMAW process. Major flaws in the welding process such as a sudden absence of shielding gas resulted in direct changes in arc signals.

The problem of stability has been approached through the study of arc signals by also varying certain characteristics of the GMAW process and observing the resulting welding results. Modenesi and Avelar (1999) investigated the influence of wire characteristics on short-

circuit GMAW stability. The authors concluded that variation of the wire diameter was the only significant factor in the onset of instability. With larger wire diameters, the spatter level may be reduced. The process current was also seen to be affected by changes in the diameter. Luksa (2006) used the monitoring of arc signals to understand the effect of artificial disturbances to the stability of the arc itself. These disturbances, either through affecting the shielding gas or through the length of the wire extension, were seen to affect the arc stability. The arc stability dictates to a great extent the quality of the resulting weld.

Another major area of research concerning the GMAW process stability relates to the formation of the metal droplets and the metal transfer mode, and their relation to stability. The metal transfer mode refers to the way in which the consumable metal electrode is melted and ultimately deposited onto the work piece. The metal transfer mode has been shown to be related to the stability of the process. Simpson (2009) presented a simplified model to relate the instability of the arc with regards to the metal transfer modes. Arc signals have also been used to sense these metal transfer modes as in the study by Johnson et al. (1991), with the ultimate goal again being the control of the GMAW process and the avoidance of instability. The droplet formation in the GMAW process has been shown to be influenced by a variety of factors and conditions. Haidar and Lowke (1997) investigated the effect of the CO₂ shielding gas on droplet formation. The effect of current on formation has been investigated by Hu and Tsai (2006). Welding velocity has also been linked to the behavior of droplets deposited onto the work piece by Li et al. (2012). Fan and Kovacevic (1999) developed an integrated model studying the droplet formation, detachment and impingement on the molten pool for the globular mode in the GMAW process.

In cases where instability or flaws are not sizable enough to yield direct visible changes in the signal, feature extracting tools need to be employed. Several time and frequency domain tools make it possible to observe certain characteristics of the signal waveforms which can provide information on the stability of the process. Statistical techniques have been used with arc signals using spectral analysis, signal variance and standard deviation (Adolfsson et al. 1999). Energy spectrum analysis, probability distributions and cyclograms were successfully extracted from arc signals to determine stability in a short circuit GMAW process by Suban, (2001). Similarly, spectral analysis was used to classify and detect major defects in the welding process in real time (Mirapeix et al. 2001). Signature analysis using power spectrum and other time-frequency tools has also yielded positive results in determining the stability of a process (Chu et al. 2004).

A promising feature extraction tool in this area is the wavelet transform. Wavelet transforms have become increasingly popular due to their ability to analyze a signal in the frequency domain for non-stationary signals (Park et al. 2008). Arc signals may be non-stationary in nature over welding runs and as such wavelet methodology may be important to consider. Wavelet transforms have been widely used in a variety of applications. The technique has proven to be very useful in detecting and investigating problems with different industrial processes. One such study was conducted by Rasoulpoor and Banejad (2013) in which the discrete wavelet transform was used to discriminate between inrush and short circuit currents. Goumas and Zervakis (2002) used the discrete wavelet transform to classify vibrations in washing machines. The vibration signals were first decomposed through the transform and then were used as extracted features in classification models to determine which signals corresponded to which type of vibrations. A similar strategy was employed in a study by Saravanan and Ramachandran (2009). In their study the discrete wavelet transform was again used to extract features from raw vibration signals. The

features were then used in a classification model using a decision tree algorithm. Feature selection was also conducted to optimize the final model. The final goal of the paper was to classify the faults of spur bevel gear boxes using these methods. Again a similar approach is seen in Muralidharan and Sugumaran (2013) in which the continuous wavelet transform for feature extraction and the decision tree algorithm for classification were implemented. The end goal was to classify faults for a centrifugal pump. Other such studies in which a wavelet transform was used to extract features in an attempt to detect differences in signals were conducted by Hong and Liang (2007) for separation of fault features, Li et. al. (2013) for rotating machinery fault diagnosis, Crossman and Guo (2003) for automotive signal fault diagnostics, in which feature selection was also conducted, and Uyar et. al. (2008) for power quality. Condition monitoring has also been explored through wavelets by Feng and Schlindwein (2009) and by Liao (2010), in which wavelet methodologies were used on acoustic emission signals in order to monitor the conditions of certain tools or parts. All these studies have incorporated wavelets in distinguishing desirable operation of a certain process from poor operation due to particular flaws or faults. The results have demonstrated the usefulness of the wavelet transform methodology as it relates to the problem at hand.

Nonetheless, wavelet transform techniques have not been as widely applied to the GMAW process. Wavelets have been used in the study of other welding processes such as laser welding (Zeng et al. 2001), pulsed metal inert gas welding (Pal et al. 2008) and gas tungsten arc welding (Wang et al. 2011). Wavelet transform methodologies have been used generally with acoustic emission signals. Processing of acoustic emission signals often relies on frequency analysis. Due to the non-stationary behavior of these signals, wavelet transforms have been implemented. Roca et al. (2009) applied wavelets to the GMAW process stability problem through acoustic emissions signals. There have been other similar studies which have used acoustic signals and

wavelets in feature extraction such as those by Lv et al. (2012) and Liao (2006). The GMAW process has also been analyzed through wavelets by Xue (2007) but they were used in image detection and not on the arc signals themselves. Recently, Zhang et al. (2014) has applied the wavelet packet transform to the Gas Tungsten Arc Welding (GTAW) process for online monitoring of welding quality. There have not been many attempts made to directly apply wavelet techniques to arc voltage and arc current signals in the GMAW process. Wavelet methodologies should be considered for GMAW monitoring in light of their efficacy in the aforementioned applications. Once features have been extracted, they can be used to predict output conditions such as stability using modeling techniques and algorithms. Techniques such as fuzzy c-means modeling, neural networks and Kohonen networks have been implemented to use features extracted from signals to monitor arc welding processes (Wu et al. 2001).

Artificial neural networks (ANN) have been widely used in developing models for the GMAW process. ANN models are very useful in analyzing non-linear processes. Furthermore, ANN models are very useful even in the absence of much knowledge of the behavior of the process. The neural network methodology was first applied to modeling the arc welding process by Andersen et al. (1990) and later by Cook et al (1995). The artificial neural network methodology was seen to be effective in predictive physical weld geometry modeling and was applied to welding process control. As a result, many efforts have been made to use ANN models to relate input parameters to a specific desired output. One such study was conducted by Ates (2007) in which the shielding gasses were modeled against desired outputs. Specifically, the hardness, tensile strength, elongation and impact strength were modeled. Similarly, other outputs related to quality such as bead geometry (Kanti and Rao, 2008), tensile strength, distortion and deposition efficiency (Pal et al. 2010) as well as weld joint strength (Pal et al. 2008) have been investigated through input arc signals with neural network modeling. The

specific problem of welding process instability due to spatter has been addressed through neural networks using acoustic emission by Roca et al. (2009) as well. The wavelet methodology based feature extraction followed by neural network modelling has been applied by Chen et al. (1999) for applications in process monitoring. However, this study was not related directly to welding processes. Saravan and Ramachandran (2010) used the discrete wavelet transform and the artificial neural network for fault diagnosis in gear boxes.

The other major intelligent modeling technique applied to this area is fuzzy modeling. As with neural networks, fuzzy logic has been extensively applied to Gas Metal Arc Welding and control. Many strategies used to intelligently control automated GMAW processes incorporate fuzzy logic such as the methodology developed by Smartt et al. (1993). In this study both neural networks and fuzzy logic were used together to utilize different welding parameters. Neuro-fuzzy systems are also gaining popularity in welding control in which parameters for the fuzzy model are trained through the ANN (Jang et al. 1997). Kim and Rhee (2001) applied fuzzy logic to arc sensor signals in an effort to control a seam tracking system. With regards to distinguishing a favorable process operation from a faulty process operation, an adaptive network based fuzzy inference system (ANFIS) was applied to classify welding defects by Zapata et al. (2010). In that study, geometrical features were extracted from image processing techniques and then classified into no-defects or into either of 5 different types of defects including slag inclusion, porosity, transversal crack and longitudinal crack. The model yielded a good classification rate and showed promise. Liao (2003) has also used fuzzy approaches in classifying welding flaw types. In both of those works, the fuzzy techniques were applied to non-destructive testing (NDT) methods and not to online monitoring. Fuzzy classification presents several advantages in weld quality due in part to the ability to retain information on how much a process is of a certain classification.

Statistical techniques have also been used to model the instability of the GMAW process. Kang and Rhee (2001) used multiple regression analysis to develop a model to estimate the amount of spatter based on short circuit transfer parameters. Arc signals were used in the model. The voltage parameter was seen to be an important factor in the onset of instability in the process. Alfaro et al. (2006) used a statistical approach to model the GMAW process relating to process monitoring.

While past research has applied feature extraction techniques and intelligent modeling approaches to the GMAW process, there are research areas which have not been explored sufficiently. For instance, as mentioned before, the wavelet methodology has not been investigated for GMAW arc signals.

CHAPTER 3- RESEARCH TOPIC AND GOAL

In many industrial situations, it may be difficult to monitor on-line any direct output condition such as welding spatter, weld strength or weld bead geometry. This paper proposes a methodology which uses only the arc sensor input signals to classify whether a process is stable or unstable based on collected online monitoring data.

The objective of this study is to devise a method to determine whether the process is operating in stable conditions or unstable conditions through the use of the input data available in real time. To this end, this thesis study intends to use a neural network and/or a fuzzy k-nearest neighbor to model extracted and selected feature data with the objective being to distinguish between stable and unstable process conditions. Additionally, this study aims to introduce wavelet transform analysis to arc signals, other than those been used in previous studies, to extract useful features. For feature selection, a wrapper approach based on ant colony optimization and fuzzy k-nearest neighbor classifier will be used.

CHAPTER 4 – EXPLORATORY DATA ANALYSIS

The welding process is fully automated and data is collected through an online data acquisition system. The process signals obtained were the arc voltage, arc current, wire feeding speed and the welding/line speed. Arc voltage, arc current and wire feeding speed signals were each collected at 1 kHz while line speed was collected at 100 Hz. The arc voltage refers to the potential difference measured across the arc, that is, between the wire and the work piece. The arc current is the measured current flowing through the wire. The wire feeding speed is the rate at which the consumable wire electrode is entering the arc. Finally, the line speed or welding speed is the rate at which the work piece is moving relative to the welding machine. If the work piece is kept stationary, the welding speed is the speed of the welding tip being pulled across the work piece. In the case involved in this study, the torch and arc are kept stationary and, as such, the welding speed is the speed at which the work piece is moving perpendicular to the torch.

Over short time intervals, the wire speed and line speed are kept constant. Fluctuations in the signals over short times (eg. less than 0.05s) may be attributed to the measuring instrument rather than due to the process itself. These measurements are taken at a very high frequency and hence may include characteristics which may not be directly related to the process but to the sensors themselves. With this in mind, changes in these two parameters can be considered over larger windows, for instance, 10 second windows. The behavior of the arc itself, however, does exhibit changes over short periods of time and should be considered for even short times.

The arc voltage and arc current are of key interest as they contain information on both the input conditions of the welding process as well as the short term behaviors of the arc. With this in mind, most features were extracted from these two signals only. Wire feeding speed and welding speed were treated primarily as input conditions and most features were not extracted from these signals.

The raw signals for both the process before and after the improvement are shown in Figure 1. As seen in the figure, the signals for both cases are visually very similar and are difficult to classify by visual observation of the raw signals alone. The wire speed signal observed on the same scale did show a prominent difference between the signals in terms of the range. The signal after improvement shows a considerably larger variation than for the signal before improvement. Upon investigating the other runs before and after improvement, this tendency was consistent. The signals after improvement had a much higher variability for this signal. This difference may be due to the measurement method or instrument rather than changes in the physical process. Based on physical considerations involving the GMAW process, differences in the wire speed are many times linked with stability issues. However, it is not expected that the variation in the signal differ this greatly at the same magnitude of the signal and that too at such small time intervals. As mentioned earlier, this signal was not used to extract most of the features which will be covered in this study.

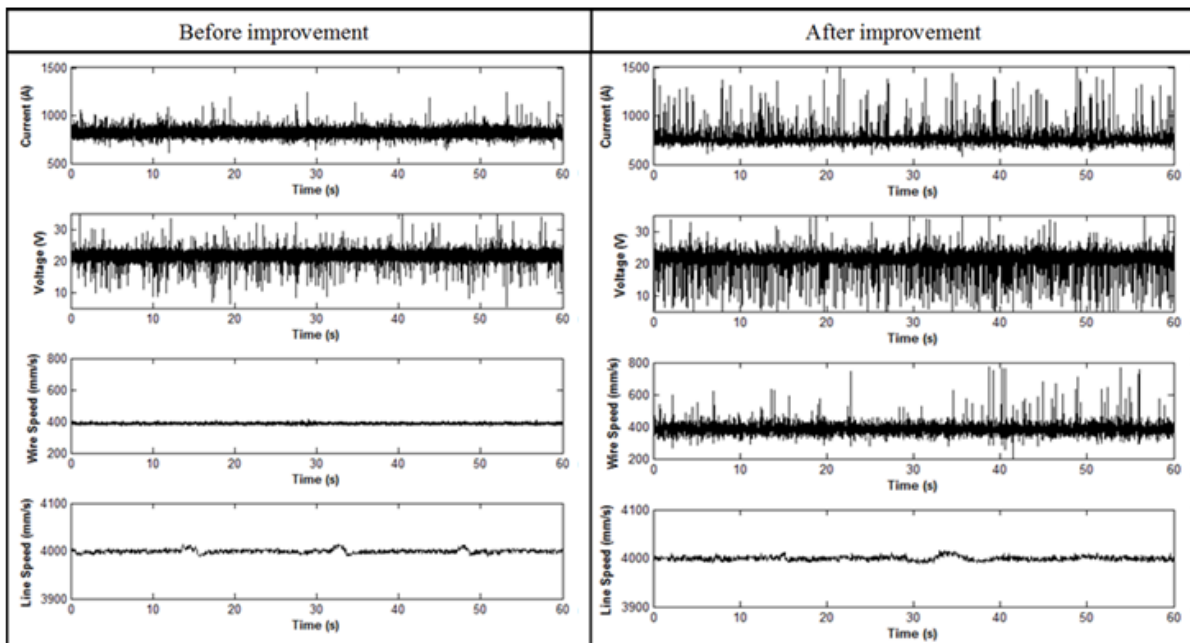


Figure 1: Process signals observed for a selected 60s window for data before and after improvement

The relationships between the signals were first investigated using scatter plots. Each variable was plotted against the others to observe the effects. This analysis was conducted on a portion of the data in which the line speed was being changed from a low to a high speed. In order to achieve the highest throughput without sacrificing quality, the line speed is often gently increased to its maximum while ensuring that the process is running smoothly. These segments of the data were ideal to observe the general relationships between the signals. In order to exclude the smallest fluctuations due to arc behavior and other factors, the average of 10 samples was used to create each point on the scatter plot. From the plots, shown in Figures 2, 3 and 4, certain relationships are evident. The data used for the scatter plots is obtained from data after process improvement as there was more varied data available from the improved process. The wire speed and the current are strongly correlated with the line speed. The reason for the strong correlation will be explained subsequently. A linear relationship is maintained between the wire speed and the line speed. The voltage for the most part is kept constant and behaves independently of the other three variables considered. These relationships were evident for both the data before and after the process improvement.

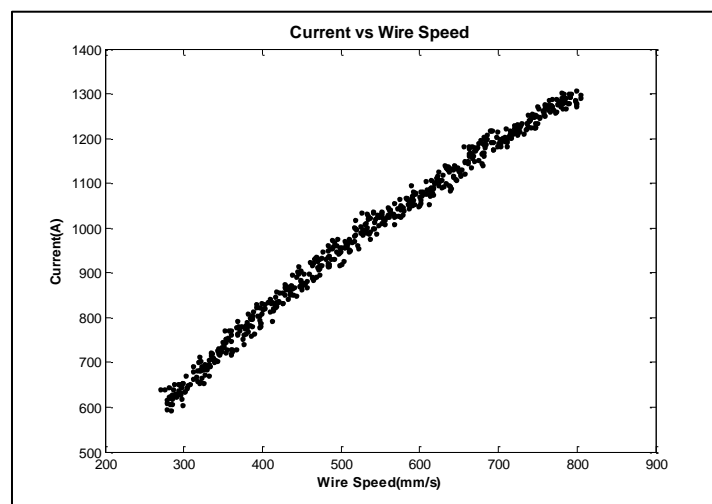


Figure 2: Relationship between the current and wire feeding speed observed over 80 s

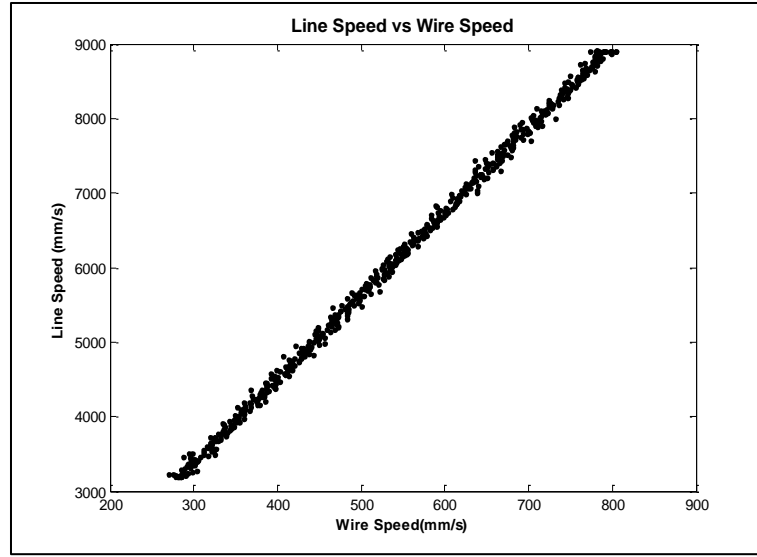


Figure 3: Relationship between the line speed and wire feeding speed observed over 80 s

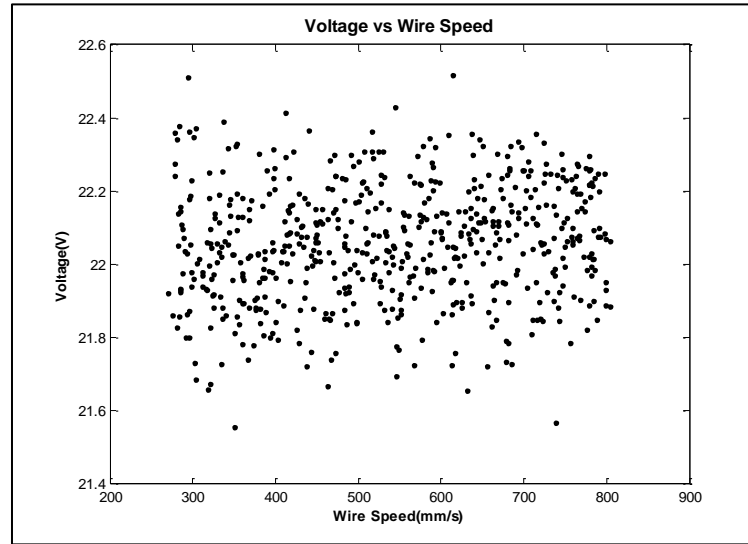


Figure 4: Relationship between the voltage and wire feeding speed observed over 80 s

The relationships between the signals were also investigated over a short time period as well. A much smaller window of 20 ms was selected from a process exhibiting some short circuiting behavior. The signal for this window is shown in Figure 5. Short circuiting areas are seen when the voltage peaks to a very high value resulting in a drop in current and vice versa. The short circuit area of the process is ideal for this analysis as it includes very large fluctuations of current and voltage.

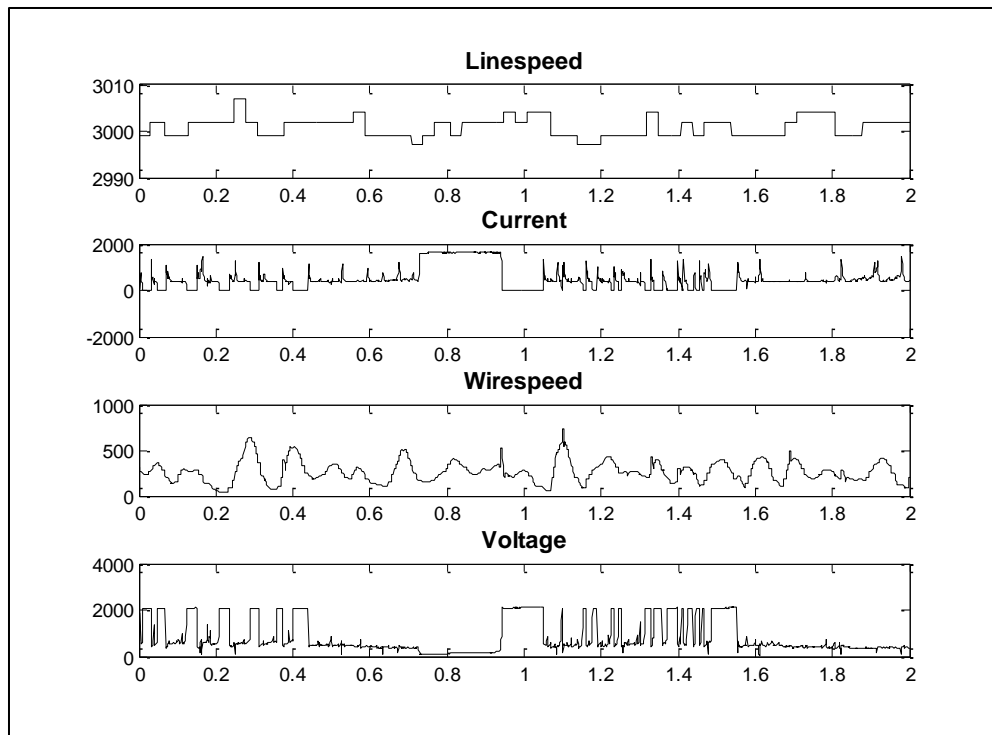


Figure 5: Signals over a time of 2 s during short circuiting

The process most closely represents a globular metal transfer process. This process is characterized by drops forming at the welding tip and detaching due to the force of gravity once drops grow to a certain critical size (Fan and Kovacevic, 1998). As the arc is formed, the electrode begins to melt. The melting rate is low enough such that drops grow to a certain size. At this size, the drops are too heavy to be attached to the electrode tip and fall to the work piece due to gravity. This process repeats itself. In spray or stream transfer, the energy is much higher and, since the melting rate is much higher, drops are not allowed to grow to this limit.

Although the process is predominantly representative of a globular process, short-circuiting transfer behavior is also present. In this transfer mode, the electrode is fed into the arc area and makes physical contact with the work piece causing a short circuit to occur. Once this short circuit is initiated, there is a large surge of energy flowing through the electrode which causes it to melt very rapidly. The molten electrode is then deposited onto the work piece. This process is

repeated. Short circuit transfer operates at low energy levels as the electrode must not melt through the arc before the short circuit is initiated as in globular or spray transfers. The short circuits are generally seen on online monitoring systems and much work has been done on analyzing signals with these short circuit patterns as in the study by Suban (2001).

As such, the metal transfer modes are very closely related with the amount of energy flowing through the system. Short circuit transfer requires the least amount of energy with globular transfer requiring more. Spray and stream transfers require the most amount of energy. These are generally the most stable and exhibit the least amount of spatter but since they require much more energy, may be infeasible to implement. The energy in the system is related to the voltage and the current and as seen before are hence related to the wire feeding speed and the line speed. There have been studies indicating that droplet rate, or the rate of deposition, affects the metal transfer mode (Heald et al. 1994), (Liu et al. 1989). The deposition is related to wire speed.

The ratio of line speed to wire speed dictates the amount of deposition per unit length of weld. This ratio is kept constant within a run to maintain a linear relationship between the two parameters and keep the weld geometry somewhat consistent. However, in some cases, this ratio has been altered between runs in an effort to optimize process conditions. The wire speed plays an important role in stability (Liu et al. 1989) and must be considered in the analysis. The voltage is controlled to be constant throughout the process. The variation in the voltage signal is hence due to the behavior of the arc and not by manual changes in the process conditions. The magnitude of the current is directly related to the operating wire speed. This is due to the requirement of more current to melt more fed wire at a constant voltage. The line speed is the main process condition which is manually varied through to maximize throughput but may result in greater chance of welding instability. In summary, current and wire speed are tethered to the line speed while the voltage is kept constant and is independent of these three parameters.

Based on the initial data observed, the proposed window size for feature extraction is 60 seconds. The 60 second window is large enough to keep the input parameter relationships mentioned earlier while still not excessively neglecting small signal variations due to the arc itself. Furthermore, the 60 second window will allow for easy extraction from welding runs which operate from anywhere between 2-30 minutes. The extraction of larger windows is infeasible due to the nature of these runs. Each window is extracted from a segment of the run in which process conditions are not being actively changed and are more or less steady.

CHAPTER 5 – METHODOLOGY

The proposed methodology exploits the difference in stability before and after process improvement to obtain a way to determine process stability. The approach intends to achieve the task at hand through three major steps.

Firstly, feature extraction techniques will be implemented on the input signals in an attempt to find characteristics of the signals which may be able to differentiate between the process data before and after the improvement. In the case that the raw signals themselves are not sufficient to distinguish between the two process conditions, certain analyses must be performed to obtain features which may be able to make this distinction. Useful features will be those which can easily highlight the differences between the two sets of data, that is, stable process and unstable process data. Several promising feature extracting techniques observed in past literature which are relevant to this process are employed. In addition, discrete wavelet transform, which has not been applied to this process before, will also be employed.

Secondly, models are developed to build a relationship between these features and the stability of the process. As nothing more specific is known about the stability of the process other than whether the process is generally stable or unstable, the stability can only be expressed as a binary variable. The relationship between the features and this stability variable will then be used to classify test data as being either from a stable or an unstable process. In this study, the used modeling techniques are the Artificial Neural Network and the Fuzzy K-nearest Neighbor Algorithm.

With a working classification model implemented, the next step is to optimize the model by selecting only the most useful features as the inputs. In the first step, feature extraction, there is no selection procedure used in deciding which features are to be used in the model. To cover a wide range of extraction techniques, all relevant feature extraction methods observed in literature

are included. The feature selection step uses the model to determine which features are the strongest distinguishers between the two process conditions. To achieve this, a meta-heuristic approach is employed to run the model with various features and combination of features and eventually obtain the best possible set of features. The final model used in this study will therefore be the reduced model obtained from this selection process. The procedures and results of these three steps are explained in greater detail in the subsequent chapters.

5.1 Feature Extraction

The features are extracted using methods reported in past research (Suban, 2001) as well as the discrete wavelet transform. Specifically, the standard deviations, energy spectra, wavelet features and other time domain factors are obtained. In all, 48 total features were extracted comprising of 4 standard deviation, 8 energy spectra, 22 wavelet energy features, and 14 other features. There is no direct disadvantage to including more features at this stage. Incorporating more features in the model may allow for a more comprehensive analysis. Each feature was obtained for the same 60 second windows selected.

5.1.1 Probability Distributions

For each window, the distributions for each signal were obtained as illustrated by Figure 6. These distributions were obtained from the same signals as the ones shown in Figure 1. Standard deviations of signals have been successfully used in past research as a tool in determining welding stability. As such, the standard deviations of each of the four signals were selected as features to be extracted from the windows. In Figure 6, the disparity in the wire speed probability distributions is evident, with the signal before improvement having a significantly smaller standard deviation than the signal after improvement. With this trend consistently present throughout the data, it is expected that this feature will have a sizable effect during the classification and selection phases of the study.

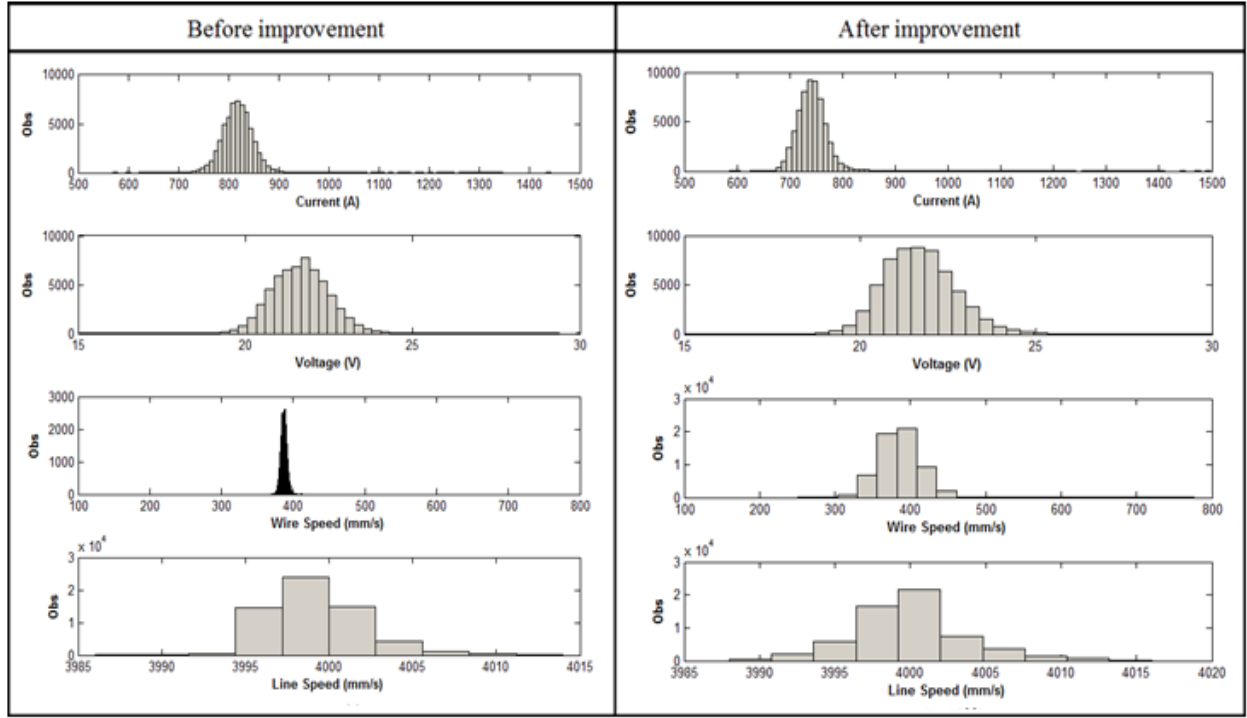


Figure 6: Probability distributions of the 4 signals in the 60s window before and after improvement

5.1.2 Energy Spectra

The signals were also observed through the frequency domain using the Fast Fourier Transform (FFT) to obtain the energy spectrum. The FFT is essentially a sinusoidal waveform for which the frequency is varied and aligned with the signal. This enables the extraction of the dominant frequencies of the signal in the window. The FFT was applied to both the current and the voltage signals.

The energy spectrum was computed from the FFT of the signal using equation (1) according to the methodology by (Suban, 2001).

$$S(\omega) = |F(\omega)|^2 \quad (1)$$

where $S(\omega)$ represents the energy spectrum and $F(\omega)$ is the fast Fourier transform.

The features extracted using this method were the frequencies and magnitudes of the four most dominant frequencies in the spectrum. To avoid clusters of these frequencies a separation distance was used to ensure a minimum spacing of 10Hz before frequencies were qualified as the dominant frequencies. The FFTs obtained for a single window corresponding to the process after improvement for both the current signal and the voltage signal are presented in Figure 7. The 4 circles plotted on each of the spectrum indicate the peaks which were extracted from the window. For each of the two signals, a total of 9 features comprised of 4 locations and 4 magnitudes of the dominant frequencies as well as the total energy were extracted using this method.

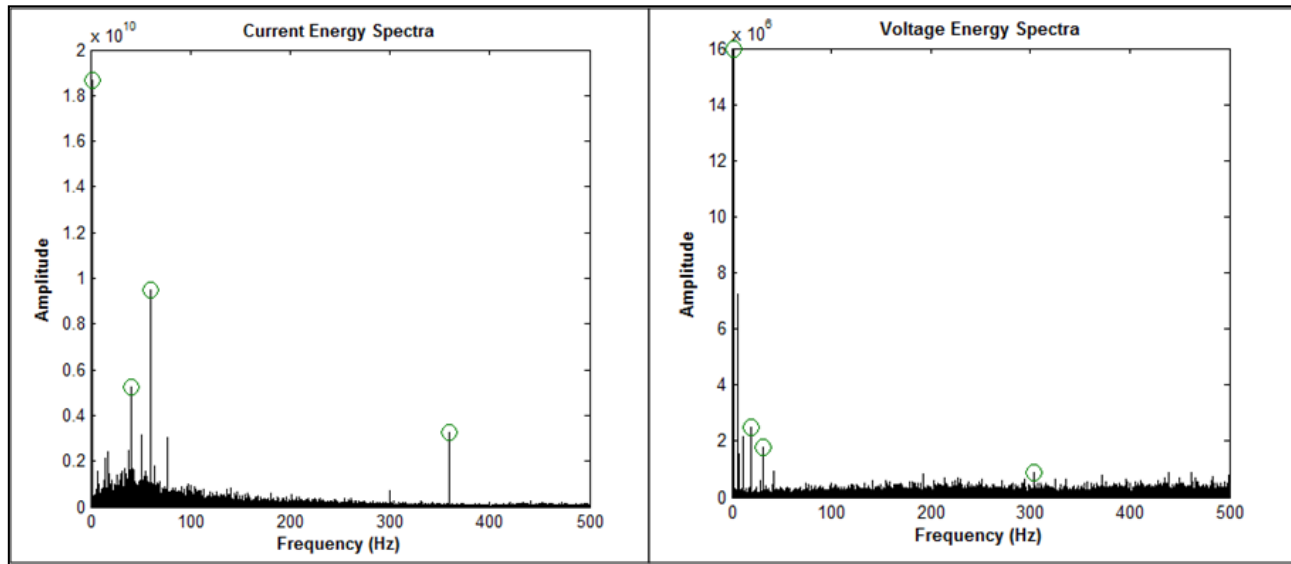


Figure 7: Energy spectra of the current and voltage signal

5.1.3 Wavelet Transform

The wavelet transform is a useful tool in analyzing the frequency domain of a signal. Wavelets are useful in analyzing non-stationary signals. The wavelet transform is more powerful in this regard than the Fast Fourier transform (FFT). One major method used to analyze non-stationary signals, in the frequency domain, was to implement the FFT over small windows

throughout the signal. By using windows, this Short Time Fourier Transform (STFT) could determine the dominant frequencies of short periods of the signal which were more so stationary. However, since a single window is used for the whole signal, there is a constant resolution. (Vetterli, 1992). Through the ability to analyze the signal with different resolutions, the wavelet transform shows more promise in non-stationary signal processing.

The continuous wavelet transform (CWT) has been replaced by the discrete wavelet transform (DWT) in many applications due to its considerably greater ease of implementation and computational simplicity. The DWT eliminates redundancies expected from the CWT and hence is preferred for this study (Liao et. al., 2007). In order to implement the DWT, first a mother wavelet must be selected. There are several families of wavelets with different advantages and hence varying preferred applications. In this study, the Daubechies wavelets are considered due to their wide use in previous research in this area. They are advantageous due to orthogonality, compact support in the time domain and computational simplicity (Saravanan and Ramachandran, 2010). Furthermore, Daubechies wavelets have been utilized due to their ability to capture the maximum amount of signal energy. Specifically, the third order Daubechies (db3) mother wavelet is used in this analysis.

The original signal is analyzed by the wavelet transform through wavelet decomposition. Each mother wavelet has an associated wavelet function and a scaling function. For the Daubechies' wavelet families, these functions can be expressed through equation 2 and equation 3 respectively (Chen et. al., 1999)

$$\varphi(t) = \sum_k^n h(k)\phi(2t - k) \quad (2)$$

$$\phi(t) = \sum_k^n l(k)\phi(2t - k) \quad (3)$$

where wavelet function is comprised of a high pass filter $h(k)$ while the scaling function is comprised of a low pass filter $l(k)$.

The original signals are convolved with the low pass filter to obtain the approximation coefficients. The signals are convolved with the high pass filter to obtain the detail coefficients. The resulting approximation and detail vectors are then downsampled resulting in vectors half the length of the original signal. These are the coefficients at decomposition level j . Convolving the downsampled approximation to the quadrature mirror filters once again yields another approximation and detail at level $j+1$ but this time at a different scale. This is illustrated in Figure 8.

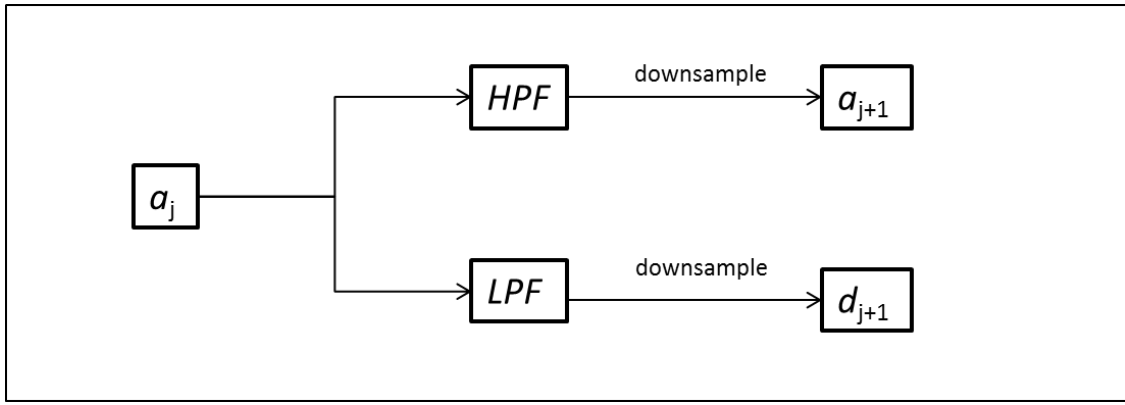


Figure 8: Wavelet decomposition at a given level j

The process is iterated until the desired number of decomposition levels is obtained. The iteration process ends with one final approximation which will not be further decomposed. This final approximation represents the signal after all the details at different frequencies have been extracted from the original signal. Hence, it shows the lowest frequency components of the signal. In other words, the general shape of the signal without the minute fluctuations can be observed at this level. Through an n -level wavelet decomposition, n details and n approximations are obtained. For the feature extraction methodology, only the details and the final approximation are considered. As such a total of $n + 1$ features are extracted. The 10-level decomposition obtained in this study is illustrated in Figure 9.

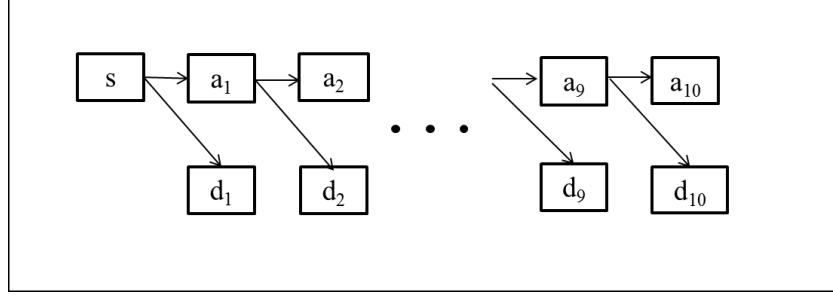


Figure 9: Wavelet decomposition process for all 10 levels of decomposition

Both the current and the voltage signals were decomposed using the DWT. A 10 level decomposition was performed using the MATLAB signal processing toolbox. The *wavedec* function was used. Since the approximations and details are downsampled by a factor of 2, decomposition can be conducted for at most $\log_2 N$ levels. With 60,000 data points available for each signal, a maximum of 15 levels could be achieved. However, the 10-level decomposition was considered and the lowest 5 frequency levels were not included. From the signals obtained through decomposition, the energy was obtained. An energy value was calculated for each level of decomposition consisting of all the details and the single approximation using equation 4.

$$E = \sum_1^N u^2 \quad (4)$$

For the wavelet application, each energy value was expressed as a percentage of the total energy. The percentages were the features extracted. Expressing the energies as percentages of the total rather than the raw values is a more sensible approach to compare the different frequency components. Keeping these normalized is useful when considering that the final intent is to determine which relative characteristics of the signal will be able to distinguish between stable and unstable data. The differences in the signal may only be detected at certain frequency levels. It also may be the case that only one frequency range reflects any differences in the signal from data before and after improvement. By covering multiple frequency ranges, the possibility of not detecting a difference reflected in the frequency domain is minimized.

By implementing the 10 level wavelet decomposition, a total of 11 features were obtained for each of the two signals. These 11 features consisted of 10 details (d1-d10) and one approximation (a10). Figure 10 presents the original current signals for the old and new data, as well as their 10 level wavelet decompositions using the db3 wavelet.

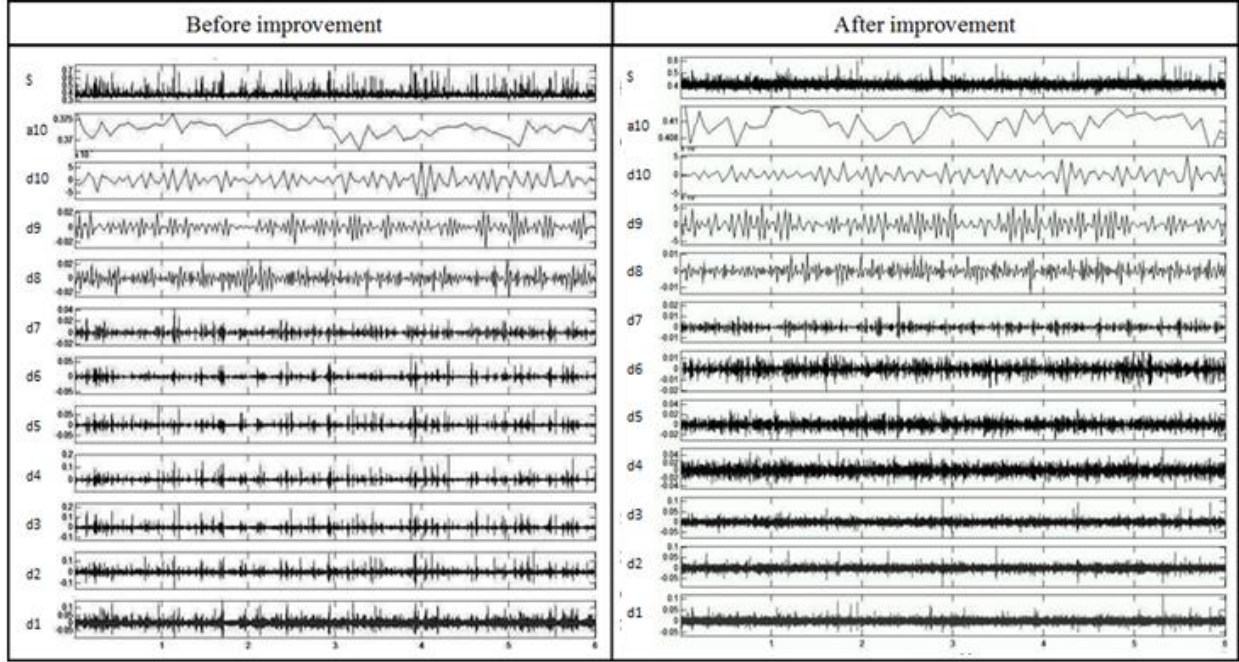


Figure 10: Wavelet decomposition results obtained from the signal before and after improvement

5.1.4 Other Time Domain Features

Due to their successful implementation with acoustic emission signals, specific time domain techniques were applied to the arc signal data to extract features. Through the methodology used in (Saini and Floyd, 1998), the RMS amplitude, absolute peak value, shape factor, crest factor, clearance factor, impulse factor and the fourth root of kurtosis were calculated for each of the two signals as shown by equations (2-8). A total of 14 features were obtained with 7 for the voltage signal and 7 for current signal. Each of these was extracted as an independent feature.

Root amplitude:

$$u_r = \left[\frac{1}{N} \sum_{n=0}^{N-1} \sqrt{|u[n]|} \right]^2 \quad (5)$$

where the u refers to the value of the sample of the signal and N refers to the total number of samples in the signal analyzed.

RMS amplitude:

$$u_{RMS} = \left[\frac{1}{N} \sum_{n=0}^{N-1} u^2[n] \right]^{1/2} \quad (6)$$

Shape factor:

$$K = \frac{u_{RMS}}{\bar{u}} \quad (7)$$

where the \bar{u} is the mean value of the signal.

Crest factor:

$$C = \frac{\hat{u}}{u_{RMS}} \quad (8)$$

where the \hat{u} is the peak value of the signal.

Clearance factor:

$$L = \frac{\hat{u}}{u_r} \quad (9)$$

Impulse factor:

$$I = \frac{\hat{u}}{\bar{u}} \quad (10)$$

Fourth root of Kurtosis:

$$\beta^{0.25} = \frac{\left[\frac{1}{N} \sum_{m=0}^{m=N-1} u^4[n] \right]^{1/4}}{u_{RMS}} \quad (11)$$

5.2 Instance Selection

All the features along with the corresponding mean input conditions were compiled into a large data table. A small portion of the table is shown in Table 1. Each row of the table represents a single window while each column aside from the first four (which represent the process condition means) represents a single feature extracted from the window. Only the first two features, namely the standard deviations are presented in the table. However, the complete table consists of 1022 rows and 30 columns. These 1022 windows were selected from both data before and data after improvement. Of these, 686 windows were selected from the data before improvement and 336 windows were selected from the data after improvement.

Table 1: Extract from the data table containing means and features extracted

Mean LS	Mean C	Mean WS	Mean V	Stdev C	Stdev V
2999.59	580.12	265.47	22.05	62.94	1.99
3999.81	748.10	347.73	22.07	55.68	1.49
2198.53	461.52	191.76	22.05	65.40	1.99
4246.28	697.83	331.77	21.61	35.47	1.54
3999.82	779.43	388.48	22.10	32.41	1.44
2049.55	418.92	176.72	21.73	101.11	2.55
4498.85	728.99	352.31	22.12	38.43	1.65
4000.60	725.57	350.55	22.09	49.96	1.34

As mentioned before, wire speed plays a very important part in the stability of the process. The values for the wire speed, line speed and the ratio were observed to be on average very different in the “after improvement” process from those in “before improvement”. After the improvement, greater line speeds could be achieved without the onset of spatter and instability. To yield greater throughput, the line speeds and wire speeds were incremented by the facility to much higher values. With such data, the model would have to use extrapolation to deal with the different input parameters used in the before and after improvement processes. Using extrapolation poses several challenges and may reduce the accuracy of the final model.

To avoid this, more efforts were placed on acquiring data from an overlapping region of input parameters. As the current is directly related to the wire speed throughout the process, the ratio and the wire speed were used to define this overlapping region. Figure 11 shows a scatter plot of the entire data with regards to the ratio (line speed over wire speed) and the wire speed. Each point represents an extracted window. The overlapping region, $R1$, is defined by box seen on the plot. The data in this region hence has similar wire speeds and line speeds. Since current is also related, the region also has similar current.

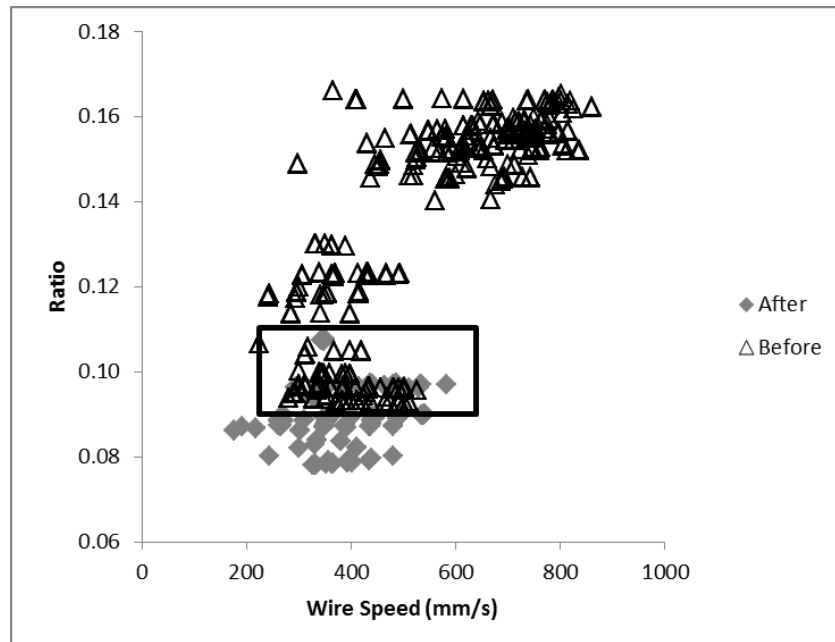


Figure 11: Plot of ratio and wire speed showing overlapping area extracted

Since the ideal overlapping area will consider all four process parameters, the voltage also had to be considered. The region, $R1$, which had been defined with respect to the other parameters, was then plotted in terms of the voltage in Figure 12. The plot showed that there were instances in which the voltage values were not similar. As such, the overlapping area had to be further refined such that the voltages were also comparable. The final region selected, $R2$, is shown by the box in Figure 12.

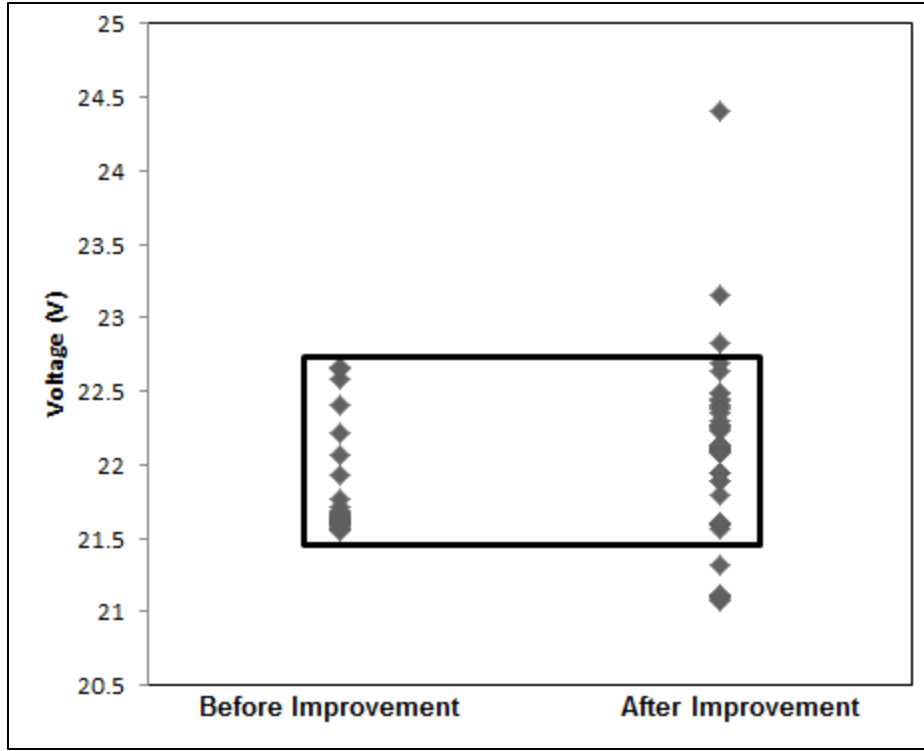


Figure 12: Selection of the final overlapping area with regards to the voltage

The final data after the definition of an overlapping area, $R2$, contained 82 windows. Of these 82 windows, 66 windows are from the data after improvement while 16 windows are from the data before improvement. Only these windows were used for the analysis in this study to remove the effects of dissimilar input parameters. There is a sizable difference in the number of windows from the data after improvement as opposed to before improvement. This poses the potential complication of analyzing unbalanced data. The effects of dealing with unbalanced data will be addressed briefly later in this study.

5.3 Classification Model

The purpose of the classification model is to use the available data in an effort to classify new data into stable or unstable condition categories. Two classification tools were considered for this task. Namely, multilayer perceptron artificial neural networks and Fuzzy-KNN models were used.

In order to distinguish between data before and after improvement, a stability variable was assigned to each window. This variable is an indicator of whether the process in a particular window represents a stable or an unstable process. A value of 1 is defined to indicate a stable process and a value of -1 is defined to indicate an unstable process. Since the data after improvement exhibits a stable process, a value of 1 is assigned to all windows originating from this data. Conversely, all windows from the data before improvement are assigned a value of -1. Both models discussed subsequently will use this same stability variable.

The unbalanced nature of the data must be considered when developing a classification model. With a much larger number of observations available from the predominantly stable condition, both the training and testing data will be biased in this regard. For unbalanced data, the classification rates are expected to be inflated. For instance, for a feature which has no relation whatsoever to the stability, the expected classification rate for a balanced case would be 50% whereas this value may be significantly inflated for the unbalanced case. This must be taken into consideration when interpreting the resulting classification rates.

5.3.1 Neural network modeling

The artificial neural network (ANN) methodology was used to develop a classification model to distinguish between the stable process data and unstable process data. The ANN was used to model the features against the stability of the process. The model was run using all the features initially but then was also run in groups of features based on the extraction methodology. Specifically, models were developed corresponding to the following groups of features: energy spectrum for the current signal, energy spectrum for the voltage signal, additional time domain characteristics for the current signal and additional time domain characteristics for the voltage signal. By analyzing groups of features as well, the classification performances can provide some insight into which features show most promise.

The neural network developed is a multi-layer perception (MLP) model with one input layer, one hidden layer, and one output layer as illustrated in Figure 13. In the network presented in the figure, four energy spectrum features are assigned as the inputs and the stability variable is assigned as the output.

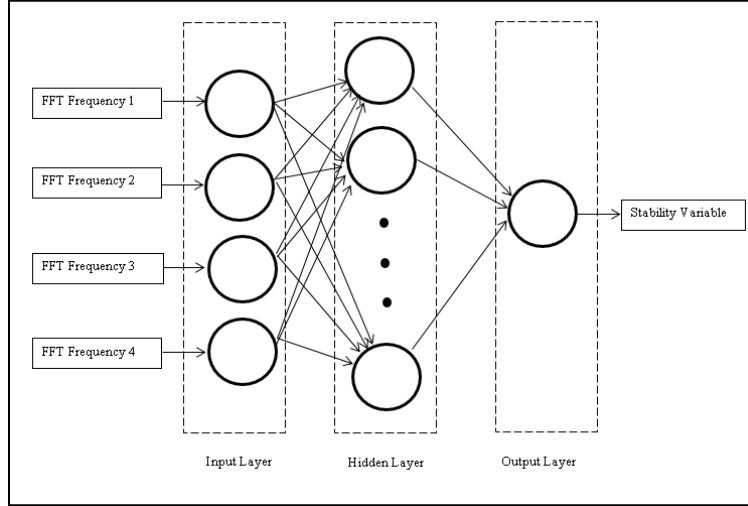


Figure 13: Multi-layer perception model structure for energy spectrum model

The input layer contains neurons which each represent a particular feature extracted from a 60 second window. The output layer contains a single neuron that corresponds to the stability variable assigned to the window. The goal of the network is to model the relationship between these nodes. This is accomplished through a hidden layer of nodes. Each input node and output node is connected to the neurons in this layer.

The hidden layer may contain any number of nodes. However, utilizing too many nodes runs the risk of over-fitting the model while too few nodes may lead to inefficient training of the model. This parameter in the final model will be determined by varying the number of nodes and obtaining the best model performance for the data being analyzed. The network architecture used in this study is a feed-forward back propagation network which is commonly seen in welding parameter models (Roca et al. 2009).

In the network, each link has a particular weight assigned. These weights are varied throughout the training process according to the back propagation algorithm which is being used. In this case, the Levenberg-Marquardt (LM) back propagation algorithm was employed as it is a fast algorithm and is widely used for input to output mapping. The LM algorithm has been successfully implemented in welding modeling networks and was shown to be the optimal algorithm when used to model input signals to weld bead width (Na et al. 2011). The hidden nodes and output nodes each contain transfer functions which determine the values which feed into the subsequent layers. Hidden nodes use the tangent sigmoid function shown in equation (12).

$$f(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)} \quad (12)$$

The output nodes use a purely linear transfer function shown in equation (13)

$$f(x) = x \quad (13)$$

The output values are determined through the weights and transfer functions and the final outputs are compared with the actual outputs observed in the training data. The resulting difference is the validation error. The model then trains by adjusting the weights for each link until a minimal validation error is achieved. The ANN model was developed using functions available in the MATLAB neural network toolbox.

The structure of the model can be varied by altering the number of nodes in this hidden layer. Too many nodes in this layer may lead to over-fitting whereas too few nodes may result in greater misclassification. The optimum number of nodes for this layer varies with regard to the data being analyzed. In order to determine the optimum for the data in this study, the number of hidden nodes was incremented from one to forty.

Of the combined dataset of 82 samples, n samples were allocated for independent testing following the k -fold cross validation scheme. The cross validation technique is widely used in modeling as it allows training and testing a data set multiple times without any redundancy in the testing data. With k being the number of folds, there will be k number of testing data segments. No observation is tested twice.

With 82 samples in this data set, each test data group would contain roughly $82/k$ observations. The remaining data was used to train and validate the model. These samples were divided such that 80% were allocated for training and the remaining 20% for validation. The validation data was used to determine the number of iterations for which the model would train. The iteration which yielded the minimum validation mean squared error was considered to be the optimal model. This validation mean squared error is calculated through equation (14).

$$MSE = \frac{1}{N} \sum (T - Y)^2 \quad (14)$$

where, T = the target value for the model and Y is the value predicted by the model at that iteration.

For the independent tests, the inputs of the $82/k$ samples were simulated through the optimally trained network. Outputs from the network yielding values below 0 predicted an unstable process while outputs above 0 predicted a stable process. These outputs were compared with the actual stability variable values and the classification rate was obtained. This process was repeated to perform k tests. Each time, a different segment of the data was used for training and for testing. At the end, the average classification rate as well as the standard deviation from the k tests was noted.

5.3.2 Fuzzy K-NN modeling

The fuzzy k-nearest neighbor methodology is considered, in part, due to the limited amount of data observations obtained from the overlapping area. The computational advantages of this algorithm also made it an attractive option to consider for this methodology. The Fuzzy K-NN method applies the fuzzy logic principles to the k-nearest-neighbor method (K-NN). The classical or crisp K-NN method could also be used in this study but including the fuzzy logic in the model has one major advantage. Fuzzy techniques allow the classification of objects to classes not just in an absolute sense but rather to what degree a member belongs to a class. In this way fuzzy can be considered much less rigid in categorizing the observations resulting in a more flexible model (Zapata et al. 2010). This is potentially advantageous in this application as there will be additional information available as to how much a new observation belongs to the unstable or stable class. With new data, this provides the ability to determine how unstable the process operates at or how much spatter is expected. For instance, if 0 indicates a ‘purely stable’ and 1 indicates a ‘purely unstable’ process, new data yielding a value of 0.51 may not be much of a concern. However, a value of 0.9 may be cause for alarm. The fuzzy membership implementation in this model allows for this distinction.

The K-NN algorithm is able to classify data into groups by computing the distance between the samples. Initially, a new observation X_n is introduced as being of an unknown class. An existing observation X_I is considered and the Euclidian distance between X_n and X_I is computed. If, based on this distance, X_n is one of the nearest neighbors of X_I then they are considered to be in the same class at that point in time. The parameter k , determines the size of the neighborhood. Only the k closest observations are used to determine the class of the new observation. The value of k , is a very important parameter to consider. Too small a value of k may lead to over fitting problems. Very large values of k , on the other hand, may lead to misclassification. The

selection of the k value for this model was accomplished in the same way as for the number of hidden nodes in the neural network. The results of this step will be discussed subsequently. With this parameter being the only major consideration, the K-NN algorithm has the added advantage of being one of the simplest classification tools.

The Fuzzy K-NN algorithm is similar to the crisp K-NN algorithm with the addition of assigning membership into a fuzzy set using equation 15 as presented by Keller et al. (1985).

$$u_i(x) = \frac{\sum_{j=1}^K u_{ij} (1/\|x-x_j\|^{2/(m-1)})}{\sum_{j=1}^K (1/\|x-x_j\|^{2/(m-1)})} \quad (15)$$

From equation 15, the Euclidean distance between the pair of nearest neighbors $\|x - x_j\|$ has an inverse relationship with the membership value $u_i(x)$. Greater distances yield lower membership values. The m parameter can be considered the strength of the fuzzy distance function (Tobin et. al., 1998). The parameter essentially dictates how much each neighbor's relative distance affects the membership value. For high assigned values of m , larger deviations of the distance have a lesser effect on their membership value, (Ghazavi and Liao, 2008). The value for m must be greater than 1 and in this case the m value was set at 2 which is a commonly used value. For an m value of 2, the membership has an inverse relationship with the distance.

The optimum k value for this model was determined by varying the value and observing the resulting classification accuracy. The full model, i.e. the model including all the features extracted, was implemented with values of k varying from 1 to 40. Since the total number of data points per feature is 82, it is expected that even higher values of k would result in greater misclassification. For each model with a different value for k , the final classification accuracy was obtained as the average from the cross validation scheme described earlier. The optimal k model was then ultimately used for the final classifications.

One possible advantage of using this methodology is that it is not as sensitive to the availability of data observations. As the model uses Euclidean distance between two neighbors to classify observations, there is no direct disadvantage in using all the features at once. Thus, it is expected that this model may perform better with all the features included than the ANN.

5.4 Feature Selection

The feature selection phase of the methodology exploits the ability of meta-heuristic algorithms in determining the best set of features. A total of 58 features were extracted using the methods described earlier. Of these 58, the optimal number of features used in the model may be anywhere from one to all 58. Intuitively, the reduction of features should reduce the computational time and complexity. One of the end goals of the overall methodology presented in this study is the application in real time monitoring in which the complexity of the algorithms used must be considered. To this end, pursuing a smaller optimal subset of features may prove to be worthwhile. Apart from simplifying the classification model, reducing the number of features translates into a simpler feature extraction phase, further streamlining the process.

Additionally, determining which features are better distinguishers between the two process-states may provide insight into the root causes of the instability. For instance, if certain wavelet frequencies are selected by the algorithm in this problem, the frequency bands corresponding to the “good” features may be investigated further. Feature extraction techniques in many processes may lead to a very large number of features of which some may be redundant or irrelevant features. Feature selection attempts to reduce this redundancy by bringing the important features related to physical characteristics of the process into clearer view (Shao et. al., 2013).

With regards to the classification model, certain features may individually be better at distinguishing between the two sets of data. However, certain features may only be capable of detecting the differences if combined with other features. Considering this possibility, it is not enough to test each feature individually and use the classification rate as the selection criteria. As mentioned earlier, two features may individually provide humble classification rates and will be excluded using this criterion. These features may, however, deliver much superior classification rates if both are included in the model and perform in unison. With 46 features in total, there is a vast number of feature sets, specifically 2^{46} , to test to determine the optimum if an exhaustive search is performed. As testing each combination is infeasible for a large n , the meta-heuristic route is preferred. By intelligently searching through the combination possibilities using a wrapper approach, the optimum set of features can be obtained with much greater ease.

Once the feature selection techniques are implemented and the optimal set of features is obtained, this set will be used in the final model. The final model will be used to determine whether new incoming data corresponds to a predominantly stable or predominantly unstable process.

5.4.1 Wrapper Approach

The four general approaches which are encountered in feature selection are the wrapper approach, the filter approach, the embedded approach and hybrid approaches (Fakhraei et. al., 2014). The wrapper and the filter approach may also be referred to as the model-based and the model-free method respectively (Chen et. al., 2013). Filter approaches use characteristics of the data to rank the features. Highest ranking features are then selected. Wrapper approaches involve the feature selection technique “wrapped around” a classification model to select optimum features. The features are selected based on their resulting classification accuracy.

This approach must use a search method to locate the optimal feature set (Tabakhi et. al., 2014). Since, for the wrapper approach the classification model has to be run repetitively for every test solution, the resulting feature selection algorithm is comparably computationally intensive especially if the wrapped model is complex. Wrapper approaches may also over-fit the subsets to a particular model. On the other hand, filter approaches provide computational efficiency and are faster but suffer in the accuracy of selection (Ulner et. al., 2011).

The feature selection process only needs to be conducted initially; the final algorithm used to analyze new data or real time data will not contain this feature selection process. Once the optimal feature subset has been determined, these features will be included in the final model. As such, there is no dire need for computational efficiency at this stage and accuracy is preferred. For this reason, the wrapper approach was chosen for the feature selection process. Although greater computational time can be tolerated at this stage, implementing the neural network in a wrapper approach would be too cumbersome. The complexity of the model has a significant effect on the computational time of the selection process since the model has to be run repeatedly to obtain candidate solutions. Hence, the simpler Fuzzy K-NN model is a more suitable option for feature selection.

Wrapper approaches have been implemented through many of the discrete optimization methods that have been developed in the recent years. Popular algorithms for this approach include the Genetic Algorithm, Particle Swarm Optimization, Simulated Annealing and the Ant Colony Optimization algorithms amongst others (Li, et. al., 2013). This study makes use of ant colony optimization, as described in more details in the following section.

5.4.2 Ant Colony Optimization

The algorithm considered for this task belongs to the Ant Colony Optimization (ACO) category of meta-heuristic techniques. The methodology was first applied to discrete optimization problems by Dorigo et al. (1999). Since then, ACO has been widely used in discrete optimization problems. To name a few, the Travelling Salesman Problem, assignment problems, job scheduling problems, routing problems and Sequential Ordering problems have been tackled by ACO.. ACO algorithms have also been extensively applied to the feature selection problem. Yan and Yuan (2004) utilized ACO in feature selection for face recognition. A recent study by Chen et al. (2013) regarding image recognition used ACO effectively for feature selection. A more general feature subset selection methodology was presented by Sivagaminathan and Ramakrishnan (2007) using ACO and artificial neural networks. Another similar approach using ANN and ACO was proposed by Kabir et al. (2012).

Ant colony optimization is based off the behaviors of ants in their colonies and their foraging habits. In nature, ants will randomly select routes towards a food source and transport food back towards the colony. During this process, they will deposit pheromones on the trails which they have travelled on. Other ants base their choice of trail on the pheromone level of the trail. Ants are more likely to select the path of greatest pheromone presence. Greater levels of pheromones on certain paths simply increase the probability of ants to follow those paths and do not prevent ants from exploring other possible paths.

The pheromones have a rate of evaporation and consequently the concentration of the pheromone on the trail will diminish with time. As such this evaporation rate plays a large part in determining the optimum path. If the rate was negligible, the first path that was chosen by the ants will continue to be the popular path. The evaporation rate prevents being trapped in local optima and allows for greater exploration. Eventually, the shortest path, or specifically the path

which requires the shortest time will be preferred as the pheromones deposited on this path will have the least time to evaporate. The ACO algorithm has been developed based loosely on this phenomenon. A specified number of ants created. The ants search for solutions through a stochastic local search policy which was used by Sivagaminathan and Ramakrishnan (2007) and then by Liao (2010).

The methods by which pheromone levels are updated vary between different ACO algorithms. In the algorithm used in this study, the pheromones are globally updated and are updated using the elitist ant strategy. In this strategy, only the best ant, i.e., the one with the best obtained solution at each iteration can update the pheromone levels. Additionally, as opposed to nature, in which the amount of pheromones deposited does not vary due to the path travelled, the algorithm determines the pheromone deposition amount based on the quality of the solution obtained by the artificial ant.

In this problem, the ants created in the meta-heuristic will attempt several combinations of features with the more effective feature combinations more appealing. The objective is to have the greatest classification rate with preferably a simpler model. Feature selection methods may include a penalty factor to favor a smaller model even at the cost of a little accuracy. However, choosing a proper penalty factor is nontrivial. The algorithm is forced to reduce the number of features by sacrificing accuracy if the penalty amount is set too high. As a result, the optimal feature subsets would not yield lower classification accuracies than the full model. It is possible, however, that the classification rate may be greater for the optimal feature subsets. Ineffective features may be reducing the classification rates depending on their effect in the Fuzzy K-NN model. Therefore, it was decided not to apply penalty factor first. In the event that after the selection process, feature subsets are still very large, penalizing feature size at the expense of classification rate may be considered.

The two variants of the algorithm considered were the ACO-S algorithm and the ACO-R algorithm. These are the sequential forward search and the random search versions of the algorithm. According to Liao (2010) in the application of feature selection for grinding wheel condition monitoring signals, the ACO-S has a superior performance as compared to the ACO-R algorithm. The optimal subsets are obtained with fewer operations. With this in mind, the ACO-S algorithm was finally selected for this study.

5.4.3 ACO-S Algorithm

The key parameters which must be taken into consideration for this algorithm are the number of ants, the number of generations, the pheromone evaporation rate and the exploitation probability factor. It must be noted that the optimal values for each of these parameters varies depending on the data analyzed. Increasing the number of ants allows for easier exploration of the solutions. Intuitively, since there are more ants searching for possible optimal solutions, more of the solution space at any given point in time can be covered. A greater number of generations increase the number of possible solutions the ants will explore.

Increasing the number of ants and the number of generations result in an increase in the number of evaluation operations conducted. This translates into greater computational load and time requirements. The other two parameters are not as directly related to computational efficiency but have other effects on the operation of the algorithm which will be briefly described subsequently.

The effects of these parameters were not tested in this study and the following values for each parameter were used: number of ants = 10, number of generations = 10, exploitation probability factor = 0.5, pheromone evaporation rate = 0.5. Increasing the number of ants and/or generations was considered initially, but due to the repeatability of the results obtained from these parameters, was deemed unnecessary.

The following pseudo-code summarizes the operation of the algorithm used by Liao (2010).

```

Load and randomize the data by observation.
Normalize the data by feature to ensure that each feature has an equal effect regardless of values.
Set num_ants=10; max_iter=10; q0 = 0.5; e = 0.5; max_s=58;
FOR r=1: Num_Runs,
P0=1,
  FOR s=1: max_s
    FOR i=1: max_iter
      Fs0 = 0, err_Fs = inf;
      FOR a=1: num_ants
        WHILE size of Fs < s
          Select feature according to equation (16)
        END while
        Compute classification error from Fuzzy KNN for selected feature subsets
      END for
      Compute the pheromone of each feature in the best feature subset according to equation (17)
      Compute the pheromone of each feature in the other subsets according to equation (18)
    END for
  END for
Keep the best feature subset for each run
END for

```

The probability of selecting a feature by the ant from the unselected feature set is given by:

$$p_{u^*(t)} = \begin{cases} 1 & \text{if } q < q_0 \text{ } u^* = \operatorname{argmax}\{\tau_u(t) \times \tau_u(t)^\beta\} \\ 0 & \text{if } q < q_0 \text{ } u^* = \operatorname{argmax}\{\tau_u(t) \times \tau_u(t)^\beta\} \\ \tau_u(t) \times \frac{\tau_u(t)^\beta}{\sum_{\forall u} [\tau_u(t) \times \tau_u(t)^\beta]} & \text{if } q \geq q_0 \end{cases} \quad (16)$$

where q is a random variable uniformly distributed over $[0,1]$. This random selection of q lends to the stochastic nature of the selection process. The parameter q_0 is initially defined and can be used to determine how exploiting or exploring the algorithm functions. For different datasets, one may be favored over the other. In the case in which $q < q_0$, exploitation is favored whereas for the case of $q \geq q_0$, exploration is favored. The β parameter is used to determine the relative importance of the pheromone. In this study, this parameter was kept as 0. For higher values of q_0 , the probability will be 0 or 1. The effect of the pheromone will have a greater impact for smaller values of q_0 .

The τ_u parameter represents the pheromone level of the feature u . This is determined at each iteration t by the following equations 17 and 18 which are used to compute the pheromone levels for a feature in the best feature subset and for a feature not in the best feature subset respectively.

$$\tau_u(t+1) = (1-e)\tau_u(t) + e/(minE(t) + \varepsilon), \quad \forall u \notin X_0(t) \quad (17)$$

$$\tau_u(t+1) = \begin{cases} (1-e)\tau_u(t) & \text{if } \tau_u(t) = \tau_u(0), \quad \forall u \notin X_0(t) \\ (1-e)\tau_u(t) + e\tau_u(0) & \text{if } \tau_u(t) \neq \tau_u(0), \quad \exists u \notin X_0(t) \end{cases} \quad (18)$$

where $minE(t)$ represents the classification error of the best feature subset.

CHAPTER 6 – RESULTS AND DISCUSSION

6.1 Classification Model

The neural network model structure was varied with regards to the number of nodes in the hidden layer as mentioned before. The number of nodes was varied from 1-40 and the resulting network was used to classify all the features together. The classification rates were obtained from each of these networks using the same k-fold cross validation scheme described earlier. Specifically, ten folds ($k=10$) were used for cross validation. These classification rates were plotted against the respective number of nodes in the hidden layer in Figure 14. From the graph, it is clear that the performance of the network reduces with the size of the hidden layer after a point. The optimum number of nodes for the network seems to be 3 nodes and as such this structure was used as the final neural network model.

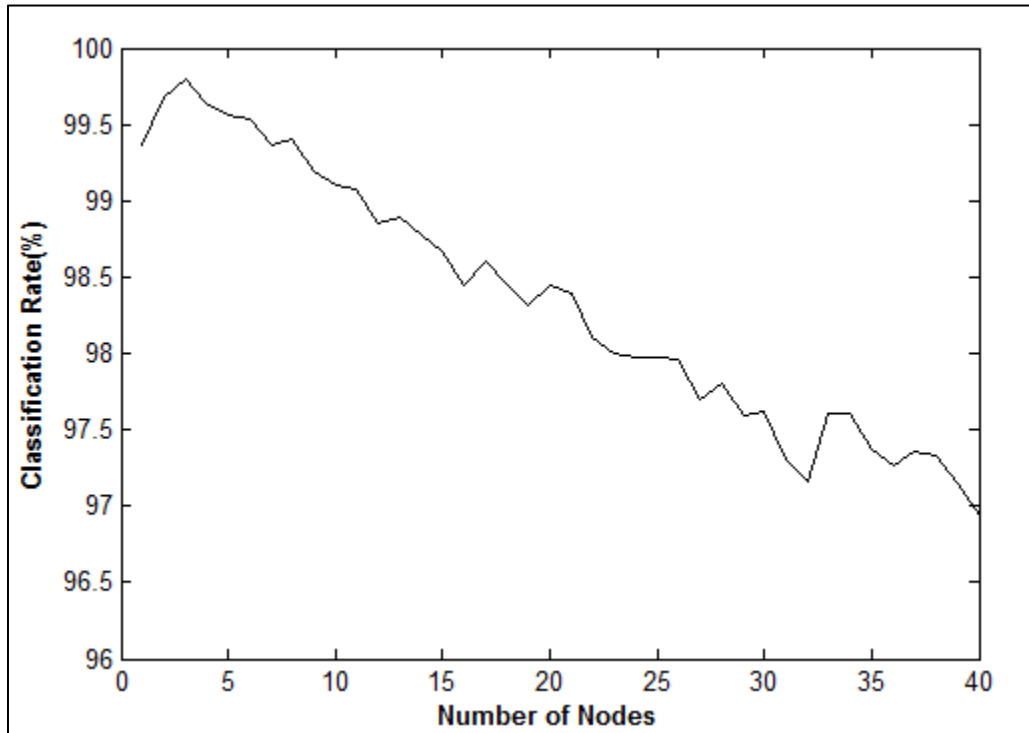


Figure 14: Resulting classification rates obtained from varying the number of nodes

Table 2 shows the classification performance for each group of features for the independent tests conducted following k-fold cross validation. The number of nodes in the hidden layer was kept at three for computing these classification rates.

Table 2: Classification performance for each feature group model with ANN

Feature Group	Classification Rate		Fp - Rate		Fn - Rate	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
Standard Deviations	100%	0%	0%	0%	0%	0%
Energy Spectrum - Current	88%	13%	6%	8%	28%	34%
Energy Spectrum - Voltage	88%	17%	9%	18%	20%	42%
Wavelet Features – Current	94%	9%	5%	8%	13%	32%
Wavelet Features - Voltage	100%	0%	0%	0%	0%	0%
Time Domain Features - Current	86%	15%	7%	10%	21%	38%
Time Domain Features – Voltage	86%	7%	3%	6%	62%	43%
All Features Included	100%	0%	0%	0%	0%	0%

From Table 2, it can be seen that the features are very effective in classifying the data. Particularly, the voltage wavelet features and the standard deviations yield a perfect classification rate. With all the features included, the classification rate remains perfect. The other ‘poorer’ features did not have a direct negative impact on the classification. The standard deviations were also included as there is some randomness in the operation of the classification model due to the randomness of the division of records. All observations are first randomized and then divided into k-folds. The way that these records are divided, especially in unbalanced data, could have an effect on the classification rates.

The false positive and false negative rates are also included to give an insight as to which group of features responsible for observations being misclassified and to what extent. False negative classifications refer to observations in the stable process conditions data being classified

as being from unstable conditions. False positive classifications are observations which are predicted as stable when in fact they originate from unstable data. The higher false negative rate than the false positive rate may be due in part to the unbalanced nature of the data. It may also be the result of the data from the process after improvement having some instabilities similar to those before improvement.

The classification accuracy was plotted against the k value of the fuzzy KNN algorithm to determine the optimum as shown in Figure 15. The results show that the optimum is anywhere between 1 and 20. As such, 1 was selected as the k value for the F-KNN model.

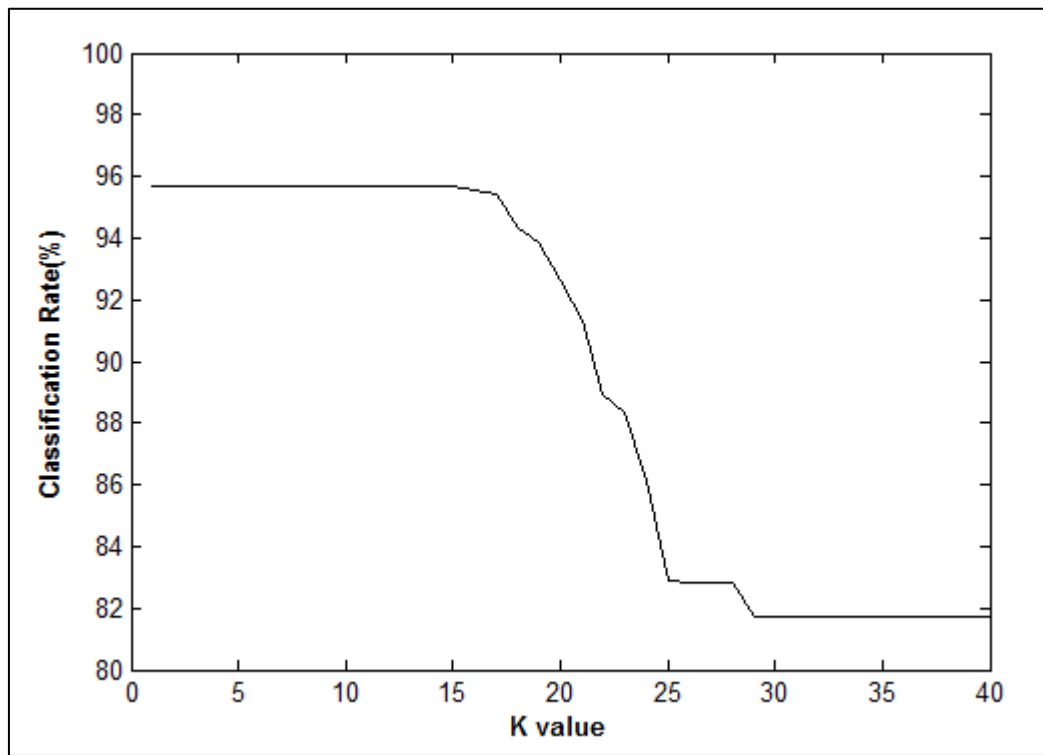


Figure 15: Resulting classification rates obtained from varying k

With k being 1, the model was run for the same sets of features as in the Neural Network model. The results are shown in Table 3. From the results, it is seen that high classification accuracies were obtained. Specifically, the highest classification was obtained from the wavelet methodology which yielded 98% classification rates for the voltage signal. Again, the standard

deviations yielded a very high classification rate of 98% due to the causes previously mentioned. The model containing all the features yielded a classification rate of 99%. The other feature groups did not yield as high classification rates and may affect the full model negatively. Through feature selection, possibly these detrimental features may be eliminated resulting in better classification.

Table 3: Classification performance for each feature group model with Fuzzy K-NN

Feature Group	Classification Rate		Fp - Rate		Fn - Rate	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
Standard Deviations	98%	5%	0%	0%	20%	42%
Energy Spectrum - Current	84%	12%	4%	10%	72%	42%
Energy Spectrum - Voltage	89%	11%	5%	7%	40%	46%
Wavelet Features – Current	88%	14%	4%	9%	38%	49%
Wavelet Features - Voltage	98%	5%	0%	0%	7%	14%
Time Domain Features - Current	85%	11%	7%	8%	34%	41%
Time Domain Features – Voltage	84%	12%	7%	10%	52%	41%
All Features Included	99%	4%	0%	0%	2%	6%

6.2 Feature Selection

The algorithm was run 100 times to explore which features were the favorites. The resulting subsets for the 100 runs were predominantly comprised of a single feature. There were optimal subsets determined which consisted of two features and there was one subset which included three features. There were 90 single feature optimal subsets obtained from the 100 runs. Figure 15 shows the distribution of these single feature subsets. From the figure it is evident that the features 38-48 were selected the most along with feature 7. These features are the voltage wavelet features and the wire feeding speed standard deviation feature respectively. The feature numbers and the extraction method for each feature can be seen in Appendix1. The selection of

the wire feeding speed standard deviation feature confirms the visual observation made earlier regarding the signal. It is clear that the voltage wavelet features are exceptional at distinguishing between the signals corresponding to the stable and the unstable processes. These single feature subsets indicate that in most the runs, each wavelet detail feature or the final approximation feature could independently detect the difference between the two data sets. The feature selection process has highlighted the utility of the wavelet transform methodology. The wavelet transform outperforms the features extracted in the time domain of the same input signals. Furthermore, the wavelet transform outperformed the Fourier transform in detecting the differences in the signals.

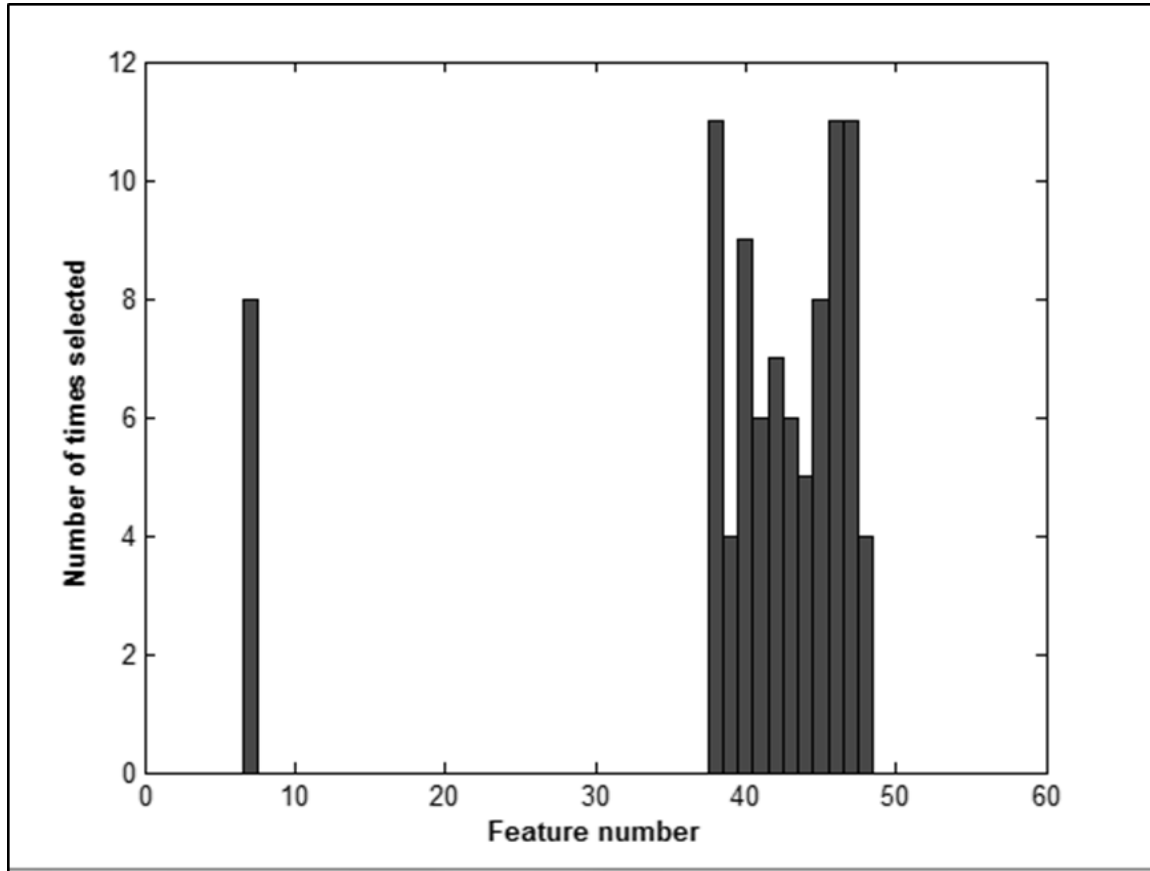


Figure 16: Single feature subsets selected by the ACO-S algorithm in 100 runs

The remaining optimal feature subsets containing multiple features are shown in Table 4. All of these feature subsets still contain the voltage wavelet features (features 38-48) or the wire speed standard deviation (feature 7) as for the single feature subsets. Since the selection criteria is very stringent in that absolutely no error in the model due to reduction is tolerated, in the event that a ‘good’ feature is selected but has even a single extra misclassification, a second or even third feature will remain in the optimal subset to ensure no further error. One characteristic of these multiple feature subsets which is worth noting is that in most subsets, the features selected are from different extraction techniques. For instance, the voltage wavelet features are coupled with either wavelet current features, FFT features or other time domain features. One possible explanation for this behavior is that features from the same extraction technique may misclassify the same data points since intrinsically they extract similar attributes of the signal. In this way, different feature extraction techniques may complement each other in the classification models.

Table 4: Multiple feature subsets obtained from the 100 runs

Two feature subsets obtained		Three feature subsets obtained		
7	36	26	40	44
7	57			
40	58			
43	2			
43	26			
43	35			
43	36			
43	58			
47	24			

Each selected voltage wavelet feature was then independently used to observe the resulting classification rates presented in Table 5. The classification rates are all remarkably high with all rates being 95% or above as expected. Each voltage wavelet feature provides comparable classification rates

Table 5: Classification performance for each feature model with Fuzzy K-NN

Feature Group	Classification Rate		Fp - Rate		Fn - Rate	
	Mean	Stdev	Mean	Stdev	Mean	Stdev
Voltage Wavelet A10	95%	6%	0%	0%	23%	34%
Voltage Wavelet D1	95%	6%	0%	0%	23%	34%
Voltage Wavelet D2	98%	5%	0%	0%	13%	32%
Voltage Wavelet D3	94%	7%	0%	0%	25%	26%
Voltage Wavelet D4	95%	6%	0%	0%	18%	24%
Voltage Wavelet D5	95%	6%	0%	0%	23%	34%
Voltage Wavelet D6	95%	6%	0%	0%	35%	47%
Voltage Wavelet D7	98%	5%	0%	0%	13%	32%
Voltage Wavelet D8	95%	6%	0%	0%	18%	24%
Voltage Wavelet D9	95%	6%	0%	0%	23%	34%
Voltage Wavelet D10	95%	6%	0%	0%	28%	42%

6.3 Ranking the features

The final wavelet features which were selected from the algorithm as single feature subsets indicated which feature extraction tool was most successful in distinguishing the old process data from the new process data. However, it may be valuable to compare these features and attempt to rank them in order of their differentiating power. As mentioned earlier, each wavelet feature describes the behavior of a signal in a different frequency range. As such, further investigation into which wavelet features are significantly different between the two datasets could provide some insight into which frequency ranges are of greatest interest with regards to distinguishing between stable and unstable conditions. The physical causes of these changes may then be investigated. Towards this goal, the voltage wavelet features were subjected to a ranking methodology. The method considered for this task was developed by Gonzales-Abril et. al. (2014). This method aims to quantify the similarity between two sets of data. This is primarily

accomplished through the use of the mean and the standard deviation of each set of data. These are used to compute the interval I^l of the data as l times of standard deviation away from the mean where l is greater than 1. The underlying principle used stems from the Chebyshev inequality which states that there is at least $(1-1/l^2)$ proportion in the interval:

$$I^l_X = (\bar{X} - l \cdot S_x, \bar{X} + l \cdot S_x) \quad (19)$$

where \bar{X} is the mean of the data and S_x is the standard deviation. The similarity between the data sets X and Y is computed through equation 20.

$$K^l_W(X, Y) = \frac{\#((X \cup Y) \cap (I^l_X \cap I^l_Y))}{\#(X \cup Y)} \cdot \frac{1}{1 + d_W(I^l_X, I^l_Y)} \quad (20)$$

where d_W is the distance between the interval obtained from data x and the interval from data y.

This is calculated through equation 21.

$$d_W(I^l_X, I^l_Y) = \sqrt{(\bar{X} - \bar{Y}, l \cdot S_x - l \cdot S_y) W \begin{pmatrix} \bar{X} - \bar{Y} \\ l \cdot S_x - l \cdot S_y \end{pmatrix}} \quad (21)$$

where W is the weight matrix which is kept at identity in this case. The first segment of equation 20 serves as a penalty factor which diminishes the similarity value for any instances in the combined data which do not exist in both intervals I^l_X and I^l_Y .

The value of l affects the length of the intervals constructed affecting both d_W and the penalty factor. This value must be selected with the nature of the data in mind. The overall objective still remains to minimize misclassification for the Fuzzy-KNN model. Hence, ideally, a feature is ‘better’ if it poses a minimal chance of misclassification. Intervals must be large enough to distinguish between features which have no overlapping region, while must be small enough such that they do not favor features with overlapping regions. The l was selected through trial and error as the smallest value which resulted a similarity value for any feature being 0. This ensured that the intervals were kept at a minimum while still giving the ability to differentiate

between features for ranking purposes. This procedure was applied to all of the voltage wavelet features with X being the data before improvement for the particular feature and Y being the data after improvement. With W kept as the identity matrix and l kept at 8, the following results were obtained as seen in Table 6.

Table 6: Similarity values obtained for each wavelet feature

Voltage Wavelet Feature	Similarity Value for $l = 8$
A10	0.300
D10	0.765
D9	0.722
D8	0.693
D7	0.525
D6	0.411
D5	0.730
D4	0.755
D3	0
D2	0.797
D1	0.806

From Table 6, the wavelet detail d3 is the best performing feature followed by a10 and d6.

This method was only suitable due to the feature selection algorithm yielding predominantly single feature subsets. If it was not clear that these wavelet features were capable of independently distinguishing between the two sets of data, this methodology could not be used since it ignores the effect of combination of features which the wrapper approach incorporates. This step is intended to provide more information on the features selected and is not by any means a substitute for the wrapper feature selection technique.

6.4 Effects of adjusting the window size

The window size used thus far in this study has been fixed at 60 seconds. However, this value was arbitrarily chosen and may not be the most optimum for this process. As mentioned before, if the current window size is too large, it may have diluted some of the important characteristics of the signal, especially those of short duration. For instance, a 60 second window, small 1-2 ms fluctuations may not have a significant effect on the features extracted. Conversely, if the currently used window size is too small, it may have omitted some characteristics of the signal occurring over larger periods of time such as trends occurring over minutes. As such, different window sizes can be implemented and the resulting classification rates observed. Additionally, adjusting the window size to smaller allows for the extraction of more records. The use of these added observations in the model may prove to be useful in classification. Since the neural network suffers in performance using very limited training data, these records allow for greater classification accuracy.

With this in mind, the window size was varied. The original results were obtained from two 60 second replicates from a 120s segment of each run. A total of 15 different window sizes were studied by dividing the 120s segment into $120/n$ windows where n varies from 1-15. The resulting classification accuracies were plotted against the number of windows used. This is presented in Figure 17. The results show that the optimum number of windows to be used for extraction is around three windows. Hence, the optimal window size is 120 divided by three or 40s. There is a sizable difference in the classification rate from one window to three windows. However, further increments in the number of windows do not result in significant changes in the model classification rate.

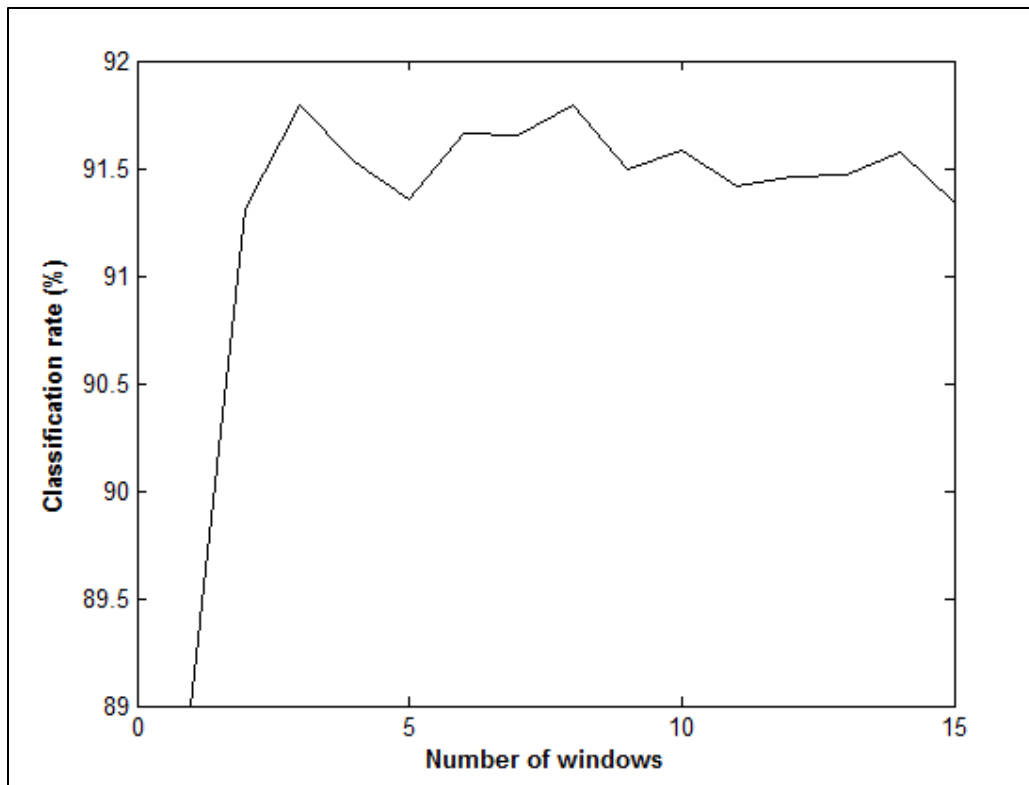


Figure 17: Effect of window size ($= 120/\text{number of windows}$) on the classification rate

Another aspect of concern regarding the window size which may be worth investigating is the effect on CPU time. This is of prime importance when dealing with online monitoring of processes in real time. Computational ease is often related to the feasibility of real time monitoring. As mentioned earlier, the feature selection and the full classification modeling phases will be independent of this real time monitoring stage. Only the selected features will be used. Furthermore, during real time monitoring, the classification model will not have to be repeatedly built. As such, only the feature extraction aspect of this process is relevant to real time monitoring concerns. It is expected that the extraction of more windows will take more computational time. However, since each window will have a shorter signal to analyze, this may not be as significant.

In order to investigate the CPU time with regard to the number of windows extracted, the feature extraction method was run for the number of windows, n , varying from one to fifteen. This was done for five different instances and the average CPU times for each n were computed and plotted as seen in Figure 18.

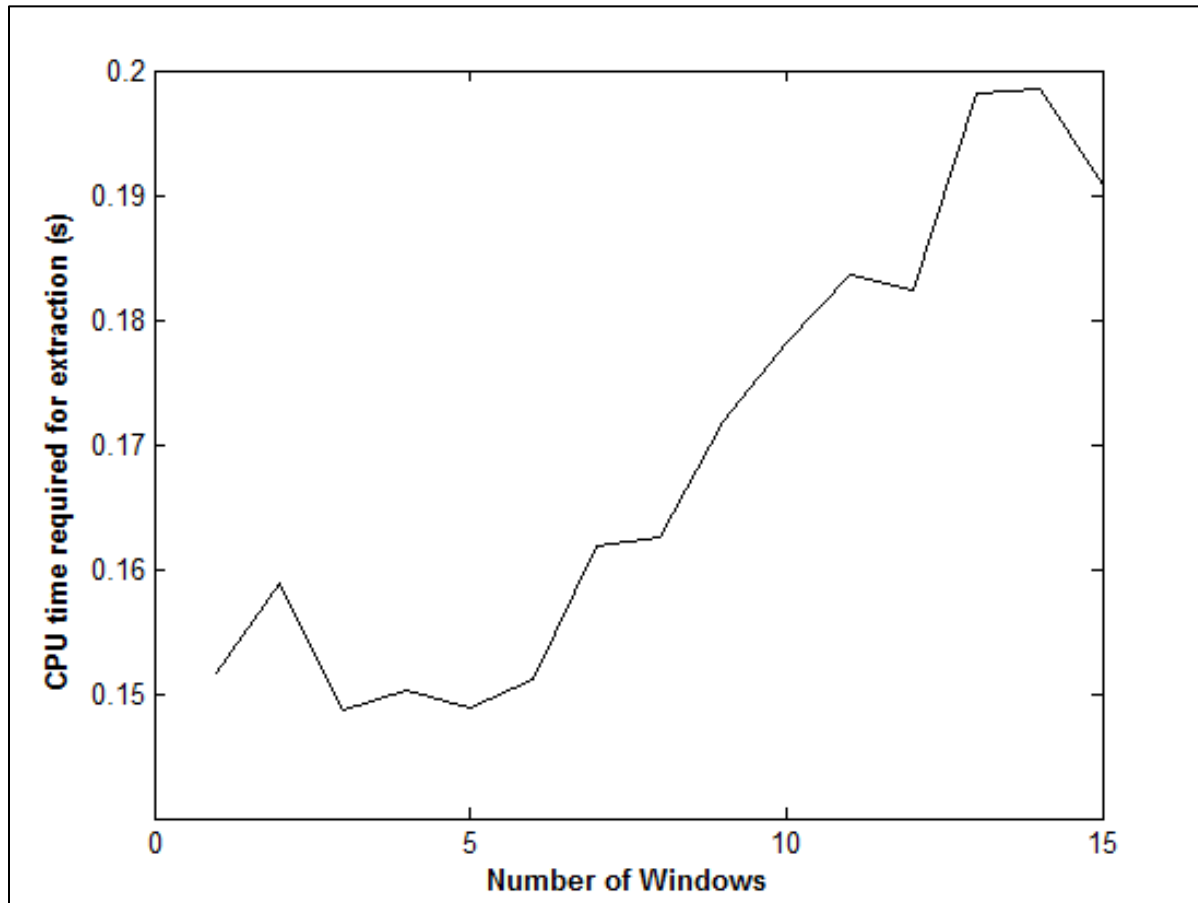


Figure 18: CPU times recorded for extracting the windows of varied sizes

There is a clear increase in CPU time with regards to the number of windows extracted. Although, the total number of data points is the same throughout and the features extracted are identical, there is more computational time required to extract a larger number of windows. Coincidentally, the minimum time observed was for $n=3$ windows which was the same for the optimum with regards to classification rate.

6.5 Application of the feature selection results to entire dataset

Due to the power of the voltage wavelet features in differentiating between the two sets of signals, it may be worthwhile to apply them to the entire dataset collected under different operating settings. To this end, these features were extracted from all the data encompassing all the process conditions available and not just from the overlapping area defined earlier. These features were then plotted for the datasets corresponding to the process before and after improvement as seen in Figure 19. From visual inspection, it is clear that some of these features may indeed be able to distinguish between stable and unstable even for different process conditions. Particularly, feature 3 or voltage wavelet detail 9 differs for the process before and after improvement.

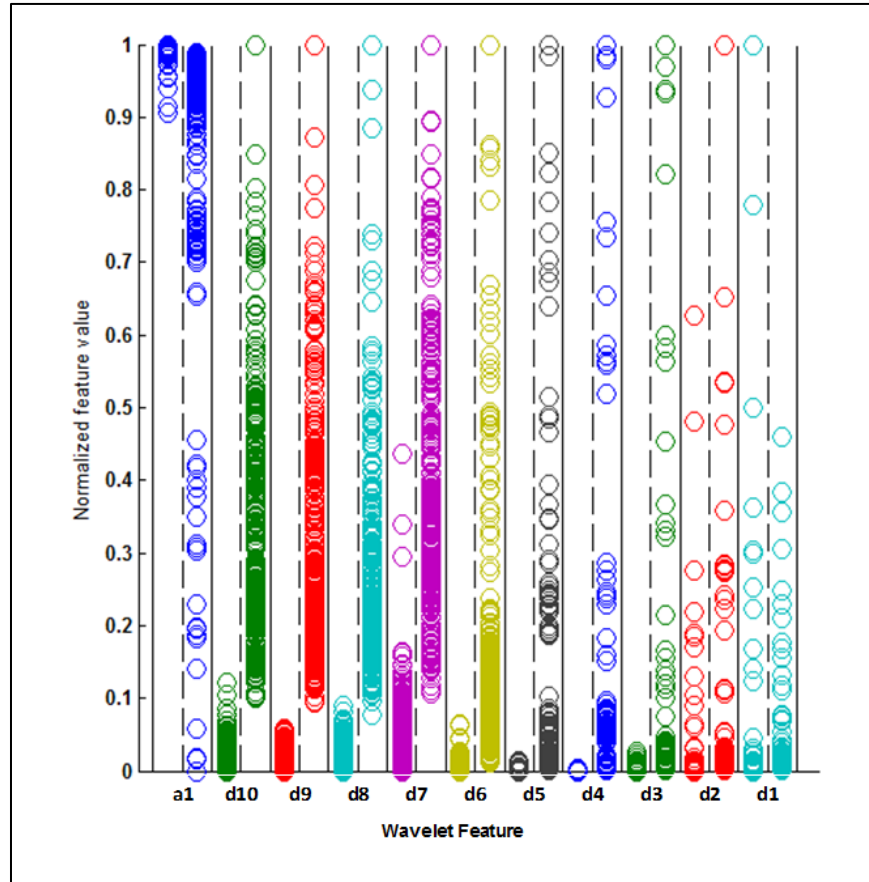


Figure 19: Normalized feature values for each wavelet voltage feature applied to entire data

This conclusion is further supported through the data similarity quantification methodology described earlier. The same technique was applied to the entire dataset to rank these voltage wavelet features. The l value was adjusted such that only one feature yielded a negligible similarity value. This value for l was 1.85. The resulting ranking can be seen in Table 7.

Table 7: Similarity values obtained for each wavelet feature for the entire data

Voltage Wavelet Feature	Similarity Value for $l = 1.85$
A10	0.119
D10	0.115
D9	0
D8	0.401
D7	0.051
D6	0.342
D5	0.245
D4	0.229
D3	0.397
D2	0.913
D1	0.966

Feature 3 or voltage wavelet detail 9 clearly has the lowest similarity value suggesting that it may have the greatest differentiating power for the data before and after.

One important factor to consider is the effect of the conditions on the feature. To investigate this, the process parameters (namely line speed, current, wire speed and voltage) were plotted against this feature. The resulting graphs, seen in figures 20-23, show that there is no significant relationship between any of the conditions and the feature considered. As such the differences in the features for the two datasets can be attributed to the differences in the processes independent of the operating settings.

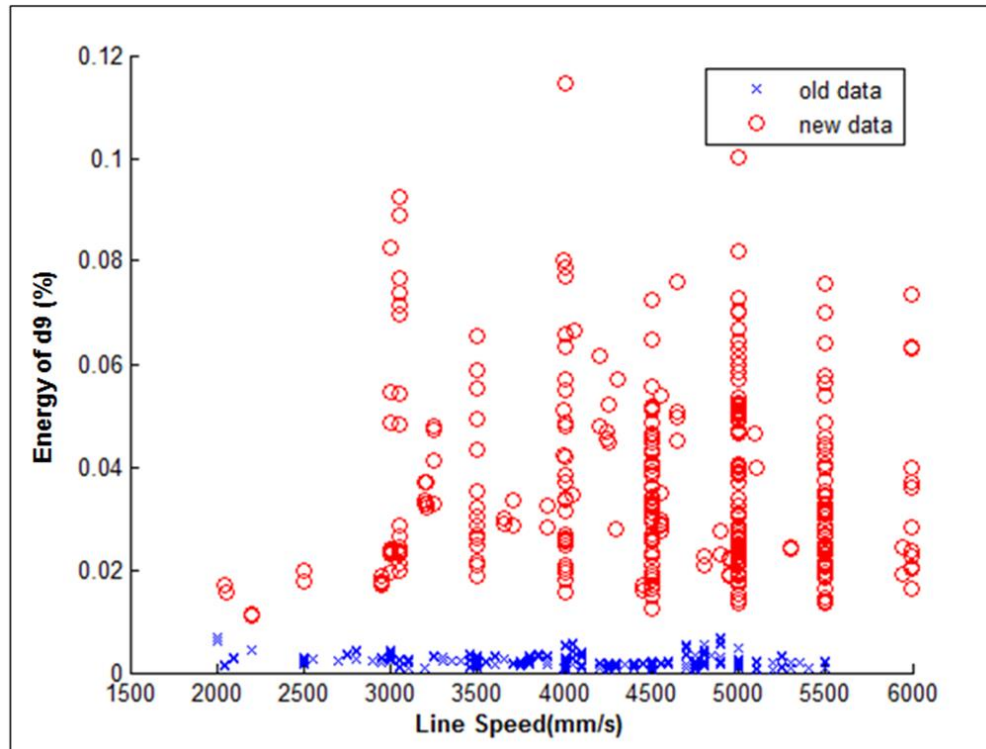


Figure 20: Relationship between the wavelet energy feature and the line speed

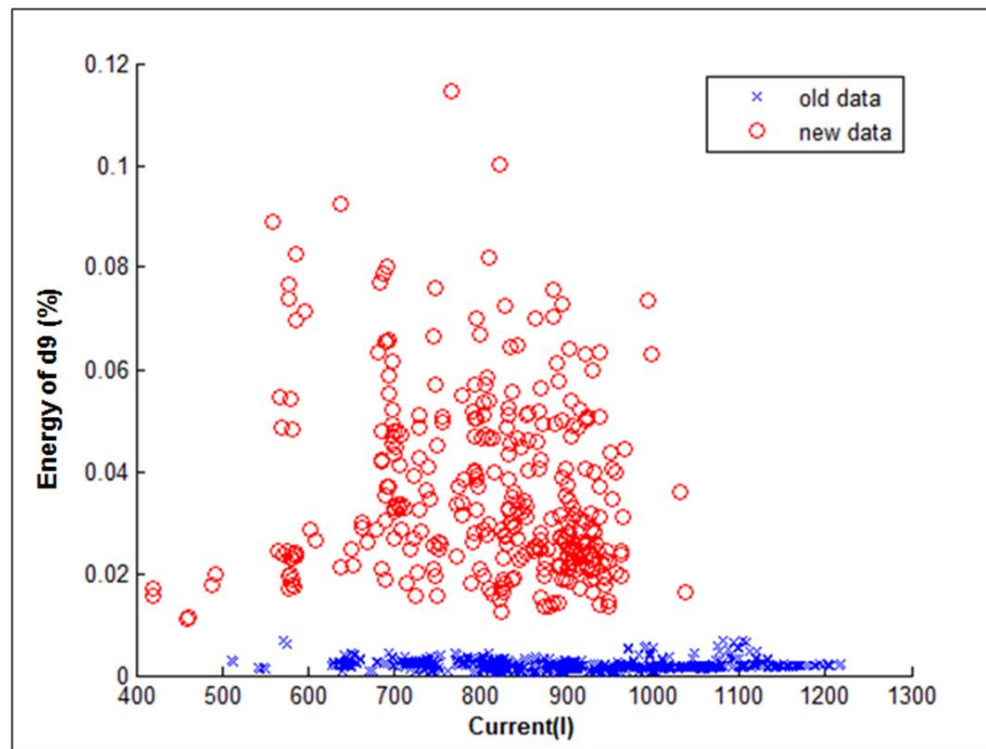


Figure 21: Relationship between the wavelet energy feature and the current

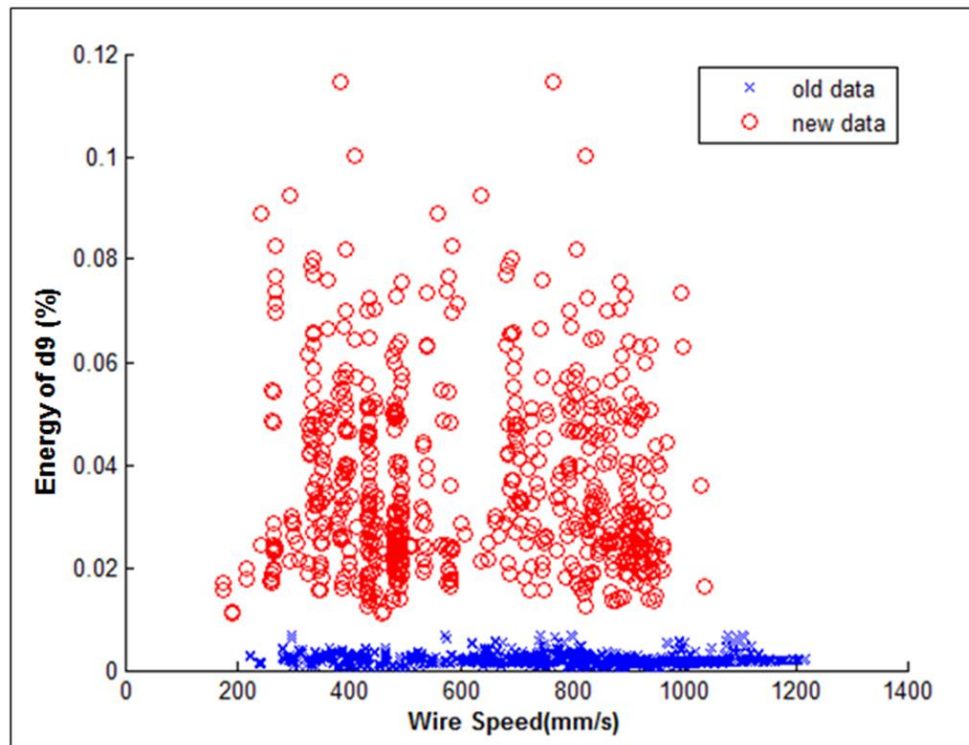


Figure 22: Relationship between the wavelet energy feature and the wire speed

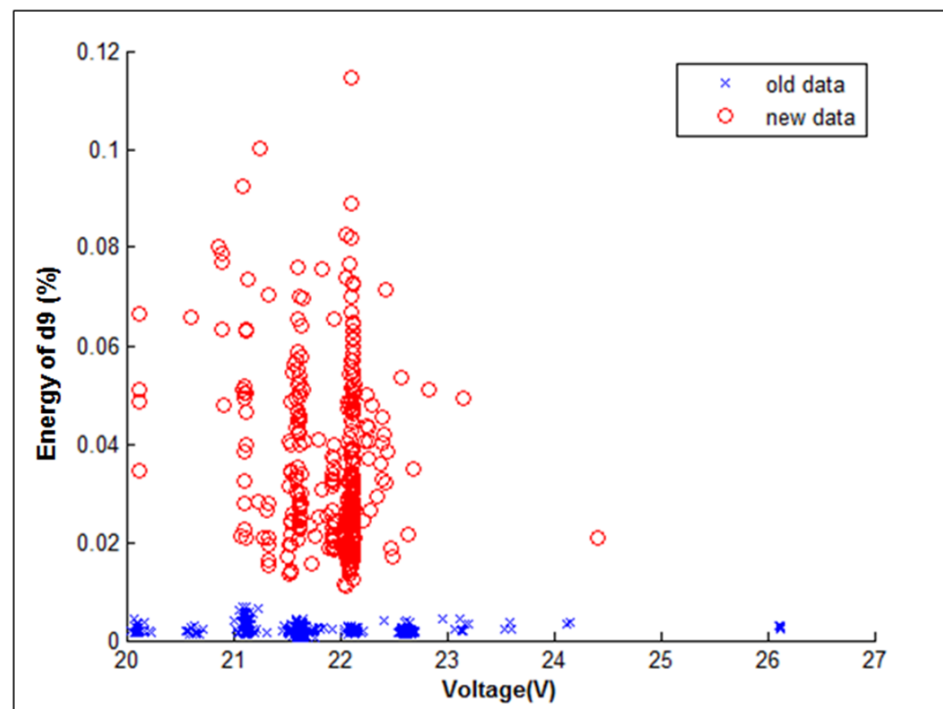


Figure 23: Relationship between the wavelet energy feature and the voltage

In all of these plots it is evident that the feature holds a different range of values depending on whether the signal represents the process before or after improvement regardless of the process condition. The implications of this finding are practically very useful in terms of process monitoring. The utility of this discovered feature lies in the fact that it can be used for different process conditions. For online conditioning monitoring, this feature can be observed in real time to predict the stability of the process for the entire range of process conditions that this process is operating at.

CHAPTER 7: SUMMARY AND CONCLUSION

On line process monitoring is a subject of great interest to industrial manufacturing processes including welding processes. Automated welding processes such as the automated Gas Metal Arc Welding can benefit greatly from analyzing and processing data through such monitoring. One major problem encountered in this GMAW process is the onset of welding spatter.

Thus, the first step of the methodology is the extraction of useful features from the raw signals which may be able to distinguish between the two process states. A total of 46 features were extracted through this phase using probability distributions, energy spectrum analysis, wavelet transform analysis and other time domain analyses.

The extracted features were then used in a classification model. The initial modeling tool utilized for this purpose was the Neural Network Model. The results of the model showed high classification rates indicating that the extracted features were indeed useful in making the distinction between the two sets of data. Particularly, the wavelet methodologies applied to the voltage signals and the probability distributions were classified with perfect accuracy. The energy spectrum analysis and the other time domain features were not as rewarding, based on the results from the model yielding classification rates in the 80-90% range.

Due to the comparably greater computational load which was characteristic of neural networks for higher dimensional problems, the more lightweight Fuzzy-KNN classifier was considered. Aside from easy implementation in wrapper approaches due to the simplicity of the algorithm, the Fuzzy-KNN methodology provides the added benefit of determining how much a new observation belongs to a certain class. This is of particular interest in online process monitoring as the classifier does not only provide a binary output such as ‘stable’ or ‘unstable’ but rather, through the membership values, gives insight on the severity of the problem.

The results obtained from this Fuzzy-KNN methodology were similar to those from the artificial neural network. The wavelet voltage features and the standard deviations were again leading in terms of the resulting classification accuracies with perfect classification. Including all the features, a classification accuracy of 99% was accomplished as compared to the 100% from the Neural Network. Although, the Neural Network yielded slightly higher classification accuracy, the Fuzzy model was ultimately chosen due to the advantages mentioned earlier.

This Fuzzy-KNN model was implemented in a wrapper approach in order to perform feature selection. The feature selection process was designed to reduce the number of features without degrading classification accuracy. The Ant Colony Optimization algorithm was used with the fuzzy model as the classifier. This algorithm was run 100 times to account for its stochastic nature. The resulting feature subsets were predominantly single feature subsets, except a few subsets, namely about 10% of the subsets, consisted of more than one feature. These results indicate that for this data, there are features which are capable of independently differentiating between the stable and unstable process data. The selected single feature subsets included all the wavelet features extracted from the voltage signals as well as the standard deviation of the wire speed signal. This further highlighted the usefulness of the wavelet methodology as applied to this data.

This methodology is not restricted to this specific problem or even the GMAW process. Rather, this feature extraction – modelling- selection process can be very useful in any manufacturing process which uses online monitoring. With regards to this specific problem, new data which is collected from the process can be processed in real time to determine whether the process operating at that time exhibits stable or unstable conditions.

Possible areas of further research with regards to this methodology may include exploration of different feature extraction techniques. There may be other ways to extract features from the signal which may compete with the wavelet features which have shown much promise in this study. With this in mind, more features can be extracted and ultimately included in the modelling and feature selection processes. Other selected features may also provide insight into the root causes of the onset of instability and spatter.

REFERENCES

1. Alfaro, A. S. C., Carvalho, G. C. and da Cunha, F. R., 2006. A statistical approach for monitoring stochastic welding processes, *Journal of Materials Processing Technology*, 175, 4-14.
2. Adolfsson S., Bahrami G., Bolmsio G. and Claesson I., 1999. On-line quality monitoring in short-circuit gas metal arc welding, *Welding Journal*, 78(2), 59s-73s,
3. Andersen K., Cook G. E., Karsai G. and Ramaswamy K., 1990. Artificial Neural Networks Applied to Arc Welding Process Modeling and Control, *IEEE Transactions on Industry Applications*, 26(5), 824-830.
4. Ates H., 2007. Prediction of gas metal arc welding parameters based on artificial neural networks, *Materials and Design*, 28, 2015-2023.
5. Campbell, S. W., Galloway, A. M. and McPherson, N. A., 2012. Artificial neural network prediction of weld geometry performed using GMAW with alternating shielding gases, *Welding Journal*, 91(6), 174s-181s.
6. Chen, B. H., Wang, X. Z., Yang, S. H. and McGreavy, C., 1999. Application of wavelets and neural networks to diagnostic system development, 1, feature extraction, *Computers and Chemical Engineering*, 23, 899-906.
7. Chen, B. L., Chen, L. and Chen, Y. X., 2013. Efficient ant colony optimization for image feature selection, *Signal Processing*, 93(6), 1566-1576.
8. Chu, Y. X., Hu S. J., Hou W. K., Wang P. S. and Marin S. P., 2004. Signature Analysis for Quality Monitoring in Short-Circuit GMAW, *Welding Journal*, 83, 335s-343s.
9. Cook, G. E., Barnett, R. J., Andersen K. and Strauss A. M., 1995. Weld Modeling and Control Using Artificial Neural Networks”, *IEEE Transactions on Industry Applications*, 31(6).
10. Crossman, J. A., Guo, H., 2003. Automotive signal fault diagnostics – Part I: Signal fault analysis, signal segmentation, feature extraction and quasi-optimal feature selection, *IEEE Transactions on Vehicular Technology*, 52(4), 1063-1075.
11. Dorigo, M., Caro, G. D., Gambardella, L. M., 1999. Ant algorithms for discrete optimization, *Artificial Life*, 5, 137-172.
12. Fan, H. G., and Kovacevic, R., 1998. Dynamic analysis of globular metal transfer in gas metal arc welding - a comparison of numerical and experimental results, *Journal of Physics D: Applied Physics*, 31(20), 2929-2941.

13. Fakhraei, S., Soltanian-Zadeh and H., Fotouhi, F., 2014. Bias and stability of single variable classifiers for feature ranking and selection, *Expert Systems with Applications*, 41, 6945-6958.
14. Feng, J., Li, L., Chen, Y., Lei, Z., Qin, H., and Li, Y., 2012. Effects of welding velocity on the impact behavior of droplets in gas metal arc welding, *Journal of Materials Processing Technology*, 212, 2163-2172.
15. Feng, Y., Schlindwein, F. S., 2009. Normalized wavelet packets quantifiers for condition monitoring, *Mechanical Systems and Signal Processing*, 23, 712-723.
16. Garcia-Allende, P. B., Mirapiex, J., Conde, O. M., Cobo, A. and Lopez-Higuera, J. M., 2009. Spectral processing technique based on feature selection and artificial neural networks for arc-welding quality monitoring, *NDT&E International*, 42, 56-63.
17. Ghazavi, S. N., and Liao, T. W., 2008. Medical data mining by fuzzy modelling with selected features, *Artificial Intelligence in Medicine*, 43, 195-206.
18. Grad L., Grum J., Polajnar I. and Slabe J. M., 2004. Feasibility study of acoustic signals for on-line monitoring in short circuit gas metal arc welding, *International Journal of Machine Tools & Manufacture*, 44, 555-561.
19. Gonzalez-Abril, L., Gavilan, J. M., and Morente, F. V., 2014. Three similarity measures between one-dimensional data sets, *Revista Colombiana de Estadística*, 37(1), 79-94.
20. Goumas, S., Zervakis, M. E. and Stavrakakis, G. S., 2002. Classification of washing machines vibration signals using discrete wavelet analysis for feature extraction, *IEEE Transactions on Instrumentation and Measurement*, 51(3), 497-508.
21. Haidar, J. and Lowke, J. J., 1997. Effect of CO₂ shielding gas on metal droplet formation in arc welding, *IEEE Transactions on Plasma Science*, 25(5), 931-936.
22. Heald, P. R., Madigan, R. B., and Siewert, T. A., Liu S., 1994. Mapping the droplet transfer modes for an ER100S-1 GMAW electrode, *Welding Journal*, 73(2), 38s-44s.
23. Hong, H. and Liang, M., 2007. Separation of fault features from a single-channel mechanical signal mixture using wavelet decomposition, *Mechanical Systems and Signal Processing*, 21, 2025-2040.
24. Hu, J., Tsai, H. L., 2006. Effects of current on droplet generation and arc plasma in gas metal arc welding, *Journal of Applied Physics*, 100
25. Huda, S., Abdollahian, M., Mammadov, M., Yearwood, J., Ahmed, S. and Sultan, I., 2014. *European Journal of Operational Research*, 237, 857-879.
26. Johnson, J. A., Carlson, N. M., Smartt, H. B., and Clark, D. E., 1991. Process control of GMAW: sensing of metal transfer mode, *Supplement to the Welding Journal*, 91s-99s.

27. Kabir, M. M., Shahjahan, M., and Murase, K., 2012. A new hybrid ant colony optimization algorithm for feature selection, *Expert Systems with Applications*, 39(3), 3747-3763.
28. Kang, M. J., Kim, Y., Ahn, S. and Rhee, S., 2003. Spatter rate estimation in the short-circuit transfer region of GMAW, *Welding Journal*, 238s-247s
29. Kang, J. and Rhee, S., 2001. The statistical models for estimating the amount of spatter in the short circuit transfer mode of GMAW, *Supplement to the Welding Journal*, 1s-8s.
30. Kanti, K. M., Rao, P. S., 2008. Prediction of bead geometry in pulsed GMA welding using back propagation neural network, *Journal of Materials Processing Technology*, 200, 300-305.
31. Keller, J. M., Gray, M. R., Givens, J. A., 1985. A fuzzy k nearest neighbor algorithm, *IEEE Transactions on Systems, Man and Cybernetics*, 15(4), 580-585.
32. Kim I. S., Son J. S., Lee S. H. and Yarlagadda P. K. D. V., 2004. Optimal design of neural networks for control in robotic arc welding, *Robotics and Computer-Integrated Manufacturing*, 20, 57-63.
33. Kim I. S., Jeong, Y. J., Son, I. J., Kim, J. Y., Kim, I. K., and Yaragadda P. K. D. V., 2003. Sensitivity analysis for process parameters influencing weld quality in robotic GMA welding process, *Journal of Materials Processing Technology*, 140, 676-681.
34. Kim, Y. and Rhee, S., 2001. Development of an arc sensor model using a fuzzy controller in gas metal arc welding, *Measurement Science and Technology*, 12, 534-541.
35. Li, P., Kong, F., He, Q. and Liu, Y., 2013. Multiscale slope feature extraction for rotating machinery fault diagnosis using wavelet analysis, *Measurement*, 46, 497-505.
36. Li, Y., Wang, G., Chen, H., Shi, H. and Qin, L., 2013. An ant colony optimization based dimension reduction method for high-dimensional datasets, *Journal of Bionic Engineering*, 10, 231-241.
37. Liao, T. W., 2003. Classification of welding flaw types with fuzzy expert systems, *Expert Systems with Applications*, 25(1), 101-111.
38. Liao, T. W., 2010, Feature extraction and selection from acoustic emission signals with an application in grinding wheel condition monitoring, *Engineering Applications of Artificial Intelligence*, 23(1), 74-84
39. Liao, T. W., Ting, C., Qu, J., Blau, P. J., 2007. A wavelet-based methodology for grinding wheel condition monitoring, *International Journal of Machine Tools and Manufacture*, 47(3-4), 580-592.

40. Liao, T. W., 2009. Improving the accuracy of computer-aided radiographic weld inspection by feature selection, *NDT&E International*, 42, 229-239.
41. Liu S., Siewert T. A. and Lau H. G., 1989. Metal Transfer in Gas Metal Arc-Welding – Droplet Rate, *Welding Journal*, 68(2), 52s-58s.
42. Luksa, K., 2006. Influence of weld imperfection on short circuit GMA welding arc stability, *Journal of Materials Processing Technology*, 175, 285-290.
43. Mirapeix J., Garcia-Allende P. B., Cobo A., Conde O.M. and Lopex-Higuera J. M., XXX Real-time arc-welding defect detection and classification with principal component analysis and artificial neural networks, *NDT&E International*, 40, 315-323.
44. Modenesi, P. J. and de Avelar R. C., 1999. The influence of small variations of wire characteristics on gas metal arc welding process stability, *Journal of Materials Processing Technology*, 86, 226-232.
45. Muralidharan, V., Sugumaran, V., 2013. Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump, *Measurement*, 46, 353-359.
46. Pal S., Pal S. K. and Samantaray A. K., 2010. Prediction of the quality of pulsed metal inert gas welding using statistical parameters of arc signals in artificial neural network, *International Journal of Computer Integrated Manufacturing*, 23(5), 453-464
47. Pal S., Pal S. K. and Samantaray A. K., 2008. Artificial neural network modeling of weld joint strength prediction of a pulsed metal inert welding process using arc signals, *Journal of Materials Processing Technology*, 202, 464-474, 2008.
48. Quinn T. P., Smith C., McCowan C. N., Blachowiak E. and Madigan R. B., 1999. Arc Sensing for Defects in Constant-Voltage Gas Metal Arc Welding, *Welding Journal*, 78(9), 322s-328s.
49. Rao Z. H., Zhou J. and Tsai H. L., 2012. Determination of equilibrium wire-feed-speeds for stable gas metal arc welding, *International Journal of Heat and Mass Transfer*, 55, 6651-6664.
50. Rasoulpoor, M., Banejad, M., A correlation based method for discrimination between inrush and short circuit currents in differential protection of pwer transformer using Discrete Wavelet Transform: theory, simulation and experimental validation, *Electrical power and Energy Systems*, 51, 168-177.
51. Roca A. S., Fals H. C., Fernandez J. B., Macias E. J. and de la Parte M. P., 2009 “Artificial neural networks and acoustic emission applied to stability analysis in gas metal arc welding”, *Science and Technology of Welding and Joining*, 14(2), 117-124.

52. Saini D. and Floyd S., 1998. An Investigation of Gas Metal Arc Welding Sound Signature for On-Line Quality Control, *Welding Journal*, 77(4), 172-179.
53. Sayyaadi, H., and Eftekharian, A. A., 2008. Modeling and intelligent control of a robotic gas metal arc welding system, *Scientia Iranica*, 15(1), 75-93.
54. Saravanan, N., Ramachandran, K. I., 2009. Fault diagnosis of spur bevel gear box using discrete wavelet features and decision tree classification, *Expert Systems with Application*, 36, 9564-9573.
55. Saravanan, N., Ramachandran, K. I., 2010. Incipient gear box fault diagnosis using discrete wavelet transform (DWT) for feature extraction and classification using artificial neural network (ANN), *Expert Systems with Applications*, 37, 4168-4181.
56. Shao, C., Paynabar, K., Kim, T. H., Jin, J. J., Hu, S. J., Spicer, J. P., Wang, H. and Abell, A. A., 2013. Feature selection for manufacturing process monitoring using cross-validation, *Journal of Manufacturing Systems*, 32, 550-555.
57. Simpson, S. W., 2009. Metal transfer instability in gas metal arc welding, *Science and Technology of Welding and Joining*, 14(4), 562-573.
58. Sivagaminathan, R. K., Ramakrishnan, S., 2007. A hybrid approach for feature subset selection using neural networks and ant colony optimization, *Expert Systems with Applications*, 33(1), 49-60.
59. Suban M., 2001. Determination of stability of MIG/MAG welding processes, *Quality and Reliability Engineering International*, 17, 345-352.
60. Tabhaki, S., Moradi, P., Akhlaghian, F., 2014. An unsupervised features selection algorithm based on ant colony optimization, *Engineering Applications of Artificial Intelligence*, 32, 112-123.
61. Tobin, K. W., Gleason, S. S., Karnowski, T. P., 1998. Adaptation of the fuzzy k-nearest neighbor classifier for manufacturing automation,
62. Unler, A., Murat, A., Chinnam, R. B., 2011. mr²PSO: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification, *Information Sciences*, 181, 4625-4641.
63. Uyar, M., Yildirim, S., Gencoglu, M. T., 2008, An effective wavelet-based feature extraction method for classification of power quality disturbance signals, *Electric Power Systems Research*, 2008, 1747-1755.
64. Vetterli, M., 1992. Wavelets and filter banks: theory and design, *IEEE Transactions on Signal Processing*, 40(9), 2207-2232

65. Wu C. S., Hu Q. X., Sun J. S., Polte T. and Rehfeldt D., 2004. Intelligent monitoring and recognition of short-circuiting gas-metal arc welding process, *Proceedings Institution of Mechanical Engineers PartB - Journal of Engineering Manufacture*, 218, 1145-1151.
66. Wu C. S., Polte T. and Rehfeldt D., 2001. A fuzzy logic system for process monitoring and quality evaluation in GMAW, *Supplement to the Welding Journal*, 33s-38s.
67. Yan, Z., Yuan, C. W., 2004. Ant colony optimization for feature selection in face recognition, *Biometric Authentication Proceedings*, 3072, 221-226.
68. Zapata, J., Vilar, R., Ruiz, R., 2010, An adaptive-network based fuzzy inference system for classification of welding defects, *NDT & E International*, 43(3), 191-199.
69. Zhang, Z., Chen, X., Chen, H., Zhong, J., and Chen, S., 2014. Online welding quality monitoring based on feature extraction of arc voltage signal, *International Journal of Advanced Manufacturing Technology*, 70, 1661-1671

APPENDIX

Table A1: Feature numbers assigned for feature selection process.

Feature #	Feature	Feature #	Feature	Feature #	Feature
1	Mean line speed	22	Location 1	43	D6
2	Mean current	23	Location 2	44	D5
3	Mean wire speed	24	Location 3	45	D4
4	Mean voltage	25	Location 4	46	D3
5	Stdev line speed	26	Energy	47	D2
6	Stdev current		Current Wavelet	48	D1
7	Stdev wire speed	27	A10		Other- Current
8	Stdev voltage	28	D10	49	Rms amplitude
	Current FFT	29	D9	50	Root amplitude
9	Magnitude 1	30	D8	51	Shape factor
10	Magnitude 2	31	D7	52	Crest factor
11	Magnitude 3	32	D6	53	Clearance factor
12	Magnitude 4	33	D5	54	Impulse factor
13	Location 1	34	D4	55	Fourth Root
14	Location 2	35	D3		Other- Voltage
15	Location 3	36	D2	56	Rms amplitude
16	Location 4	37	D1	57	Root amplitude
17	Energy		Voltage Wavelet	58	Shape factor
	Voltage FFT	38	A10	59	Crest factor
18	Magnitude 1	39	D10	60	Clearance factor
19	Magnitude 2	40	D9	61	Impulse factor
20	Magnitude 3	41	D8	62	Fourth Root
21	Magnitude 4	42	D7		

VITA

Mr. Aniruddha Joshi completed his high school education in the Charlestown Secondary School, Saint Kitts and Nevis in June 2006. He then obtained an Associate degree in Natural Sciences from the Charlestown Secondary Sixth Form College, Saint Kitts and Nevis. He then went on to obtain his bachelor of science in mechanical engineering in May, 2012 from the Georgia Institute of Technology, Atlanta, GA. He is a candidate for the Master of Science degree in Industrial Engineering in December, 2014. Aniruddha's research interests are in process monitoring and metaheuristics.